Replicability and Generalizability of Posttraumatic Stress Disorder (PTSD) Networks: A Cross-Cultural Multisite Study of PTSD Symptoms in Four Trauma Patient Samples

Fried, E.I.; Eidhof, M.B.; Palic, S.; Costantini, G.; Huisman-Van Dijk, H.M.; Bockting, C.L.H.; ... ; Karstoft, K.-I.

Replicability and Generalizability of Posttraumatic Stress Disorder (PTSD) Networks: A Cross-Cultural Multisite Study of PTSD Symptoms in Four Trauma Patient Samples

Eiko I. Fried [1], Marloes B. Eidhof[2], Sabina Palic[3],
Giulio Costantini[4], Hilde M. Huisman-van Dijk[5],
Claudi L. H. Bockting[2,6], Iris Engelhard[5,6], Cherie Armour[7],
Anni B. S. Nielsen[8,9], and Karen-Inge Karstoft[8]

[1]Department of Psychology, University of Amsterdam, Amsterdam, The Netherlands; [2]Arq Psychotrauma Expert Group Diemen/Oegstgeest, The Netherlands; [3]Competence Center for Transcultural Psychiatry, Mental Health Center Ballerup, Copenhagen, Denmark; [4]Department of Psychology, University of Milan-Bicocca, Milan, Italy; [5]Altrecht Academic Anxiety Centre, Utrecht, The Netherlands; [6]Department of Clinical Psychology, Utrecht University, The Netherlands; [7]Psychology Research Institute, Ulster University, Coleraine Campus, Northern Ireland; [8]Research and Knowledge Center, The Danish Veteran Center, Ringsted, Denmark; and [9]The Research Unit and Section of General Practice, Institute of Public Health, University of Copenhagen, Denmark

Abstract

The growing literature conceptualizing mental disorders like posttraumatic stress disorder (PTSD) as networks of interacting symptoms faces three key challenges. Prior studies predominantly used (a) small samples with low power for precise estimation, (b) nonclinical samples, and (c) single samples. This renders network structures in clinical data, and the extent to which networks replicate across data sets, unknown. To overcome these limitations, the present cross-cultural multisite study estimated regularized partial correlation networks of 16 PTSD symptoms across four data sets of traumatized patients receiving treatment for PTSD (total $N = 2{,}782$). Despite differences in culture, trauma type, and severity of the samples, considerable similarities emerged, with moderate to high correlations between symptom profiles (0.43–0.82), network structures (0.62–0.74), and centrality estimates (0.63–0.75). We discuss the importance of future replicability efforts to improve clinical psychological science and provide code, model output, and correlation matrices to make the results of this article fully reproducible.

Keywords

posttraumatic stress disorder, replicability, network modeling, generalizability, open materials

The network approach to psychopathology has received increasing attention and recognition in recent years and has been used to study a plethora of mental disorders, including depressive disorders (Fried, Epskamp, Nesse, Tuerlinckx, & Borsboom, 2016), generalized anxiety disorder (Beard et al., 2016), posttraumatic stress disorder (PTSD; McNally et al., 2015), eating disorders (Forbush, Siew, & Vitevitch, 2016), and psychosis (Isvoranu, Borsboom, van Os, & Guloksuz, 2016; see Fried et al., 2017, for a review of the empirical literature and important concepts). The core idea is that problems (often symptoms) cluster in specific constellations (syndromes)

because they are associated in causal webs and vicious circles (Borsboom & Cramer, 2013). In other words, a mental disorder like depression arises not from one central brain dysfunction that gives rise to all symptoms but from problems interacting in dynamic systems that can be hard to escape. Clinical network theory has been explained in detail in several recent publications

Corresponding Author:
Eiko I. Fried, University of Amsterdam, Department of Psychology, Niewe Achtergracht 129-b, Amsterdam 1001NK, Netherlands
E-mail: eiko.fried@gmail.com

(Borsboom, 2017; Cramer, Waldorp, van der Maas, & Borsboom, 2010; Hayes, Yasinski, Ben Barnes, & Bockting, 2015; McNally, 2016), and we will refrain from reiterating it here in more detail.

These theoretical insights have led to the recent development of psychometric models, often referred to by the umbrella term *network models* (Bringmann et al., 2013; Epskamp & Fried, 2017; van Borkulo et al., 2014). The aim of these models is to estimate network structures of psychological variables from between-subjects or within-subjects data. Network models are largely exploratory and data driven, and although they use tools such as regularization to avoid overfitting data (Friedman, Hastie, & Tibshirani, 2008), it is presently unclear whether the findings from these network models replicate across different data sets, a question especially relevant considering the recent attention to replicability in psychology (Open Science Collaboration, 2015; Tackett et al., 2017). Quite appropriately, researchers working with network models have questioned whether we are about to face a replicability crisis in this newly developing field and what can be done to avoid it (Epskamp, Borsboom, & Fried, 2017; Forbes, Wright, Markon, & Krueger, 2017a; Fried & Cramer, 2017). One important way forward is to routinely test and report the precision of statistical parameters derived from network models, which can safeguard against overinterpretation. To give one example, if edges A—B and C—D have weights (connection strengths) of 0.7 and 0.5, respectively, it is unclear whether the first edge is meaningfully or significantly stronger than the second without testing the precision of these parameters, for example, by obtaining confidence intervals around the parameter estimates via bootstrapping routines (Epskamp, Borsboom, & Fried, 2017; Fried & Cramer, 2017). A second way forward is to empirically test whether network structures generalize across different data sets. This article, for the first time, investigates this question across four clinical data sets of patients receiving treatment for PTSD.

Network models were implemented only recently in the field of PTSD research (McNally et al., 2015) and have been used in at least 11 articles since (Afzali, Sunderland, Batterham, et al., 2017; Afzali, Sunderland, Teesson, et al., 2017; Armour et al., 2016; Birkeland & Heir, 2017; Bryant et al., 2017; Frewen, Schmittmann, Bringmann, & Borsboom, 2013; Knefel, Tran, & Lueger-Schuster, 2016; Mitchell et al., 2017; Spiller et al., 2017; Sullivan, Smith, Lewis, & Jones, 2016). Overall, we identify three specific challenges in the prior literature of PTSD symptom networks that we aim to address in this article. First, PTSD network studies estimated networks in one sample only, and it is unclear how the results generalize across populations of different cultures,

trauma types, or different levels of clinical severity (Marsella, Matthew, Friedman, Gerrity, & Scurfield, 1996). Replicability efforts across PTSD data sets are especially relevant given that trauma reactions are heterogeneous, and different trauma types are associated with different symptom profiles (Kelley, Weathers, McDevitt-Murphy, Eakin, & Flood, 2009). Forbes, Wright, Markon, and Krueger (2017b) argued recently that the results of network models estimated in single PTSD data sets do not seem to be highly consistent across studies. This aligns well with the fact that factor-analytic methods applied to PTSD symptom data have yielded conflicting results about the optimal factor structure (Armour, Műllerová, & Elhai, 2015).[1] This apparent lack of consistent results strongly warrants replicability investigations. Second, only a few PTSD network articles featured large samples (Bryant et al., 2017; Mitchell et al., 2017); most publications are based on comparably small populations with only about 200 participants (Armour et al., 2016; Birkeland & Heir, 2017; Knefel et al., 2016; Spiller et al., 2017). Given that network models require the estimation of many parameters and that these models need considerable power to reliably detect small coefficients (Epskamp, Borsboom, & Fried, 2017; Epskamp & Fried, 2017), investigations of larger data sets are needed. Third, studies have applied network models to PTSD symptom data in only community (e.g., Afzali, Sunderland, Batterham, et al., 2017; Afzali, Sunderland, Teesson, et al., 2017; Sullivan et al., 2016) or subclinical/mixed samples (e.g., Armour et al., 2016; Bryant et al., 2017; Knefel et al., 2016; McNally et al., 2015; Mitchell et al., 2017). The network structure in clinical samples—arguably the most relevant level of observation if we take network theory seriously—is presently unknown. All three limitations are acknowledged as crucial challenges in the recent literature (Bryant et al., 2017; Epskamp, Borsboom, & Fried, 2017; Fried & Cramer, 2017).

Our cross-cultural, multisite study addresses these three points by investigating the similarities and differences of network structures of PTSD symptoms in four moderate to large data sets of traumatized patients receiving treatment for PTSD with different index trauma types, including civilian-, refugee-, combat-, postwar offspring-, and professional duty-related trauma. The article makes two additional contributions. First, we use a recently developed network estimation technique to jointly estimate symptom networks across the four data sets based on the fused graphical lasso (FGL) that can lead to a more accurate estimation of network structures than estimating networks individually (Costantini et al., 2017; Danaher, Wang, & Witten, 2014). The FGL improves network estimates by exploiting similarities among different groups where such similarities

Table 1. Demographics of Four Clinical Samples of Traumatized Patients Receiving Treatment

	Sample 1	Sample 2	Sample 3	Sample 4
Description	Treatment-seeking patients	Treatment-seeking patients	Treatment-seeking soldiers	Treatment-seeking refugees
Data collected in	Netherlands	Netherlands	Denmark	Denmark
Patients (*N*)	526	365	926	965
Age [*M* (range)]	47 (17–74)	35.6 (18–61)	36.2 (21–76)	NA (18–79)
Females (%)	35.9	72.1	5.2	42
(Probable) PTSD diagnosis (%)	67.7	100	59.3	100
Symptom severity [*M* (*SD*)]	2.76 (0.66)	2.70 (0.58)	2.36 (0.77)	3.21 (0.42)

Note: NA = not available; PTSD = posttraumatic stress disorder. Mean age of participants in Sample 4 is unknown; patients were not asked about specific age (only age categories). The majority of patients (41%) were in the age range of 40 to 49 years.

emerge; otherwise, networks are estimated independently. Second, although we cannot share the data sets themselves, the Supplemental Material available online includes all R code, model output, descriptive statistics, and correlation matrices of the data sets. Network models (like factor models) in ordinal and continuous data can be estimated on the basis of the correlation matrix and do not require the raw data as model input, which makes the results of this article fully reproducible and allows for future investigations of the clinical data sets we analyzed here.

Method

Participants

We analyzed four samples of traumatized patients receiving treatment (total *N* = 2,782). Characteristics of the four samples are depicted in Table 1; details can be found in the Supplemental Material. All patients were assessed for the presence of PTSD symptoms before treatment or within 3 months of starting treatment.

The first sample consisted of 526 traumatized patients who were enrolled at Arq, a Dutch mental health center specializing in treatment of patients with severe psychopathology and a history of complex psychotraumatology like war, persecution, profession-related traumatic events, and other complex traumatic events. The sample consisted of refugees (36%), patients traumatized during the course of professional duty (soldiers and police officers; 24%), postwar generation offspring (24%), and victims of other human violence (16%). All patients were assessed with the Harvard Trauma Questionnaire (HTQ; Mollica et al., 1992), a self-report instrument, as part of the routine diagnostic procedure for all patients who were referred for treatment. Using a cutoff score of 2.5 (average HTQ symptom on a scale of 1–4), 67.7% of this sample had probable PTSD. Data were collected between 2001 and 2015.

Sample 2 consisted of 365 traumatized patients from Altrecht Academic Anxiety Centre, a Dutch outpatient clinic specializing in treatment of anxiety and related disorders encompassing various trauma types. As part of the routine diagnostic procedure, all patients filled out the Posttraumatic Stress Symptom Scale Self-Report (PSS-SR; Foa, Cashman, Jaycox, & Perry, 1997) and were interviewed by a trained clinician using the Structured Clinical Interview for *DSM–IV* Axis I Disorders (SCID). All participants included in this study had a diagnosis of PTSD according to the SCID. Data collection took place between 2008 and 2016.

The third sample consisted of 926 previously deployed Danish soldiers receiving treatment for deployment-related psychopathology at the Military Psychology Clinic within the Danish Defense or were referred for treatment at specialized psychiatric clinics or psychologists in private practice. As part of the routine diagnostic procedure for all treatment-seeking patients, self-reported PTSD symptoms were assessed using the Civilian version of the PTSD checklist (PCL-C; Weathers, Litz, Herman, Huska, & Keane, 1993). Using the PCL-C cutoff score of 44, validated as the best cutoff for probable diagnosis in an independent sample of Danish soldiers (Karstoft, Andersen, Bertelsen, & Madsen, 2014), 59.3% of the patients had probable PTSD. Data were collected between 2014 and 2016.

Sample 4 consisted of 956 refugees with a permanent residence in Denmark. The data were pooled from the Danish Database on Refugees With Trauma (Carlsson, Sonne, & Silove, 2014), run by the Competence Centre for Transcultural Psychiatry (part of the Danish mental health system, situated in Copenhagen). Patients underwent routine clinical assessment for the presence of psychological disorders according to the ICD-10 diagnostic criteria and filled out the HTQ. All patients were diagnosed with PTSD, and approximately 30% suffered from persistent trauma-related psychotic symptoms. Fifty-two percent came from different Arabic-speaking

countries (Palestine, Iraq, Lebanon, Syria), 13% were from Iran, 13% from the countries in the former Yugoslavia, 11% from Afghanistan, and the remaining 10% from other countries such as Chechnya and Somalia.

Missing data

Overall, there were very few missing values on the 16 PTSD symptoms: 9, 2, 3, and 37 for Data Sets 1 through 4, respectively. We excluded these participants when necessary, for example, when estimating the symptom means and standard deviations. For the network analysis, we retained all participants and estimated the correlations among symptoms on the basis of pairwise complete observations.

Measures

To assess the presence and severity of PTSD symptoms according to the fourth edition of the *Diagnostic and Statistical Manual of Mental Disorders* (*DSM–IV*; American Psychiatric Association, 1994), the 16-item HTQ (Samples 1 and 4), 17-item PSS-SR (Sample 2), and 17-item PCL-C (Sample 3) were used. All are widely used self-report instruments with Likert-type scales ranging from 1 (*not at all*) to 4 (*extremely*) for the HTQ, 1 (*not at all*) to 4 (*very much/almost always*; rescaled from original 0–3 range to fit the other scales) for the PSS-SR, and 1 (*not at all*) to 5 (*extremely*) for the PCL-C. The HTQ and PSS-SR assess symptoms during the past week, whereas the PCL-C measures symptoms during the past month. The difference in number of items is explained by the fact that the PCL-C and PSS-SR—in contrast to the HTQ—assess physiological and emotional reactivity separately. To allow for a comparison of the measures, we combined these two items of the PCL-C and PSS-SR to fit the format of the HTQ (highest score on either of these two symptoms was used for the analysis). Finally, to compare the means across scales, we rescaled the PCL-C to the same range as the other instruments (1–4).

We computed internal consistency (Cronbach's α on the basis of the polychoric correlations) and composite reliability (on the basis of the factor loadings of unidimensional confirmatory factor analysis models). Reliability scores for the questionnaires used in Samples 1 through 4 (HTQ, PSS-SR, PCL-C, and HTQ), calculated via Cronbach's α and composite reliability, were 0.91/0.92, 0.89/0.87, 0.94/0.93, and 0.85/0.80, respectively.

Statistical analyses

We conducted the analysis in four steps: network estimation, network inference, network stability, and network comparison. All analyses were carried out in R version 3.3.1 in RStudio 1.0.136. We used the R package *qgraph* (Epskamp, Cramer, Waldorp, Schmittmann, & Borsboom, 2012) to visualize all networks. All exact version numbers of all R packages used are documented in the Supplemental Material.

Network estimation. The present state-of-the-art for ordinal or continuous data is to estimate a Gaussian Graphical Model (GGM; Lauritzen, 1996), a network in which edges connecting symptoms represent estimates of partial correlations. In the GGM, edges can be understood as conditional dependence relations among symptoms: If two symptoms are connected in the resulting graph, they are dependent after controlling for all other symptoms. If no edge emerges, symptoms are conditionally independent. GGMs are typically estimated using the graphical lasso, a method that employs regularization to avoid estimating spurious edges (Friedman et al., 2008). This method maximizes a penalized log-likelihood, a log-likelihood function plus a penalty term that depends on network density (the number and the strength of edges). A tuning parameter (λ_1) allows regulating the importance of the density penalty. Larger values of λ_1 yield sparser networks (i.e., with fewer and weaker edges), whereas smaller values yield denser networks. Because it is unknown whether the true network is sparse or dense, the value of λ_1 is typically selected empirically, using *k*-fold cross-validation (i.e., train and validate the model on different parts of the data and choose the value of λ_1 that results in the best prediction) or information criteria, such as the extended Bayesian information criterion (Epskamp & Fried, 2017). Using the graphical lasso to estimate a GGM improves network estimates and leads to a sparse network that describes the data parsimoniously. The method has been used and explained in numerous recent articles, and an accessible tutorial article on GGM estimation and regularization is available elsewhere (Epskamp & Fried, 2017).

In our case, we aimed to accurately estimate the GGMs in four groups of individuals. If the true networks in these samples were the same, the most accurate network would be obtained by estimating a single GGM using graphical lasso on the full data set. However, this strategy would ignore differences across groups. Conversely, estimating four individual networks would allow detecting such differences but would result in poorer estimates if the networks were the same (because of lower power in each data set compared with the full data). The FGL (Danaher et al., 2014) is a recent extension of graphical lasso that allows estimating multiple GGMs jointly. Like the graphical lasso, FGL includes a penalty on density, regulated by the tuning parameter λ_1. Unlike the graphical lasso, the FGL also includes a

penalty on differences among corresponding edge weights in networks computed in different samples, regulated by a tuning parameter λ_2. Large values of λ_2 yield very similar networks in which edges are estimated by exploiting all samples together; small values allow network estimates to differ; and λ_2 of zero means that networks are estimated independently. Because it is unknown whether the true networks are similar or different, a principled way of choosing both λ_1 and λ_2 is through k-fold cross-validation. Overall, FGL improves network estimates by exploiting similarities among groups. If this does not improve model fit, the k-fold cross-validation procedure selects a value of the λ_2 parameter equal or very close to zero, in which case separate GGMs are estimated via the graphical lasso. As a result of this strategy, the FGL neither masks differences nor inflates similarities across groups. The FGL has been used successfully to compute gene expression networks in cancer and healthy samples (Danaher et al., 2014), to estimate networks of situational experience in different counries (Costantini & Perugini, 2017), and to examine borderline personality disorder symptom networks in patients and healthy individuals (Richetin, Preti, Costantini, & De Panfilis, 2017; for a tutorial on the FGL, see Costantini et al., 2017).

In this article, we estimated networks in the four samples using FGL and selected optimal values of λ_1 and λ_2 parameters via k-fold cross-validation, as implemented in the R package *EstimateGroupNetwork* (Costantini & Epskamp, 2017). Because FGL yields generally better network estimates (Danaher et al., 2014), we report this joint estimation as the main model in the article. However, because networks in the literature have been typically estimated using graphical lasso, the Supplemental Material contains results obtained by estimating networks individually. Additionally, we report the results of a different method for selecting the tuning parameters for FGL via information criteria instead of cross-validation. Both methods led to nearly identical results to those reported here.

Network inference. We computed centrality indices for the four jointly estimated networks. Whereas previous articles have often investigated three different measures of centrality—betweenness (i.e., the number of times a specific node lies between two other nodes on their shortest connecting edge), closeness (i.e., the inverse of the summed length of all shortest edges between a node and all other nodes), and node strength (i.e., the sum of all edges of a given node to all other nodes; McNally et al., 2015)—recent investigations have shown that betweenness and closeness are often not reliably estimated (Epskamp, Borsboom, & Fried, 2017). This was also the case in our analyses, and we thus focus on node strength

in the remainder of the article, while reporting betweenness and closeness in the Supplemental Material.

We also estimated shared variance of each node with all of its neighbors, which is referred to as *predictability* in the literature (Haslbeck & Fried, 2017), using the R package *mgm*. In contrast to centrality, which is a *relative* metric of how interconnected a node is, predictability provides us with an *absolute* measure of interconnectedness. It can also be understood as an *upper bound* to controllability: If we assume that all connections go toward this node, predictability quantifies how much influence we can have on this node by intervening on all its neighbors.

Network stability. We used the R package *bootnet* to investigate the stability of the networks. Stability estimation has only recently been developed (Epskamp, Borsboom, & Fried, 2017) and is not yet worked out for jointly estimated networks. We instead examined the stability of the individual networks, and results thus provide a *lower bound* for stability in the jointly estimated networks. We bootstrapped 95% confidence intervals around the edge weights, estimated the correlation-stability coefficient for centrality metrics (ranging from 0–1; values above 0.25 imply moderate stability, above 0.5 strong stability), and computed the edge-weights difference test and the centrality difference test for each network. These methods are described in detail elsewhere (Epskamp, Borsboom, & Fried, 2017), and results are described in the Supplemental Material.

Network comparison. Finally, we compared the four networks in several aspects. First, we correlated the edge weights across networks, which provides a coefficient of similarity (Borsboom et al., 2017; Rhemtulla et al., 2016). Second, we formally tested whether the networks differed from each other in their network structures via the R package *NetworkComparisonTest* (NCT; van Borkulo et al., 2017). To this end, we started with an omnibus test for each pair of networks to investigate whether all edges were exactly identical; this was followed by post hoc tests to quantify how many of the 120 edges differed across each pair of networks. For this post hoc test, the NCT uses the Holm-Bonferroni method to correct for multiple testing.[2] Third, we used NCT to test whether global strength estimates (the sum of all absolute edge values for each network) differed across networks. Fourth, we estimated and visualized the cross-sample network. We averaged the edge weights across the networks instead of estimating a network by pooling all participants into one data set because the latter would have given more weight to the larger data sets (note that our procedure likely leads to a less sparse network compared with an estimated network on all data sets, because an

Table 2. Overview of the 16 Posttraumatic Stress Disorder Symptoms (Including Means and Standard Deviations) From Four Clinical Samples of Traumatized Patients Receiving Treatment

No.	Symptom	Short codes	M (SD)			
			Data 1	Data 2	Data 3	Data 4
1	Intrusions	Intr	3.10 (0.91)	3.15 (0.86)	2.41 (1.08)	3.43 (0.68)
2	Nightmares	Nightm	2.66 (1.12)	2.45 (1.02)	1.97 (1.15)	3.33 (0.76)
3	Flashbacks	Flash	2.61 (1.08)	2.60 (0.97)	1.76 (1.04)	3.19 (0.81)
4	Physiological/psychological reactivity	React	2.84 (1.01)	2.86 (0.89)	2.35 (1.11)	3.47 (0.66)
5	Avoidance of thoughts	AvThought	2.78 (1.03)	2.85 (1.10)	2.18 (1.17)	3.05 (0.95)
6	Avoidance of situations	AvSit	2.74 (1.10)	2.38 (1.09)	1.85 (1.14)	3.26 (0.87)
7	Amnesia	Amnes	1.96 (0.99)	2.26 (1.09)	1.90 (1.14)	2.34 (1.13)
8	Disinterest in activities	Disint	2.77 (0.97)	2.76 (1.08)	2.62 (1.13)	3.18 (0.87)
9	Feeling detached	Detach	2.80 (0.94)	2.52 (1.02)	2.70 (1.11)	3.24 (0.87)
10	Emotional numbing	EmNumb	2.39 (1.05)	2.43 (1.05)	2.47 (1.12)	2.56 (1.07)
11	Foreshortened future	ShortFut	2.79 (1.07)	2.95 (1.07)	2.07 (1.17)	3.42 (0.84)
12	Sleep problems	Sleep	3.08 (1.00)	3.20 (0.97)	2.98 (1.14)	3.51 (0.67)
13	Irritability	Irrit	2.65 (0.98)	2.45 (0.90)	2.68 (1.07)	3.30 (0.80)
14	Concentration problems	Conc	3.12 (0.88)	2.87 (0.91)	2.86 (1.02)	3.48 (0.70)
15	Hypervigilance	Hyperv	3.05 (0.94)	2.81 (0.99)	2.72 (1.17)	3.21 (0.87)
16	Startle response	Startl	2.91 (0.94)	2.61 (0.93)	2.26 (1.18)	3.31 (0.83)

Note: To allow comparison of means and standard deviations across data sets, all questionnaires were rescaled to have a range of 1 to 4.

edge is nonzero in our case if it is nonzero in any of the data sets). Fifth, to highlight similarities and differences across the four individual networks, we estimated a cross-sample variability network in which each edge (e.g., between A—B) depicts the standard deviation of this edge A—B across the four networks, similar to a previous article (Rhemtulla et al., 2016); strong edges imply greater variability.

Availability of data and materials

The analyses performed were not formally preregistered. The analytic code for all analyses performed in this study is available in the Supplemental Material, along with Supplemental figures, tables, correlation matrices, and other R objects that allow researchers to reproduce our results (e.g., symptom means and standard variations, covariance matrices among symptoms, network parameters, results of all stability analyses). The original data cannot be shared because of restrictions of the clinical institutions in which they were gathered; further details on how to apply for the data are available from the corresponding author on request.

Results

Descriptive statistics

Samples differed in average symptom endorsement: Patients in Data Set 4 had the most severe symptomatology,

followed by Data Set 1, Data Set 2, and Data Set 3 (see Table 1). Except for the comparison of Data Set 1 with Data Set 2, $t(840.15) = 1.62$, $p = .11$, Bayes factor = 0.26,[3] all other differences between the severity scores were highly significant (t values between 8.51 and 29.29, degrees of freedom between 518.03 and 1417.3, all p values < 2.2×10^{-16}, all Bayes factors > 4.7×10^{13}). Table 2 lists all symptoms and short codes.

The lower variability of the symptoms in Data Set 4 was also reflected in the variability of the individual symptoms (see Table 2), and there were indications of a ceiling effect in Data Set 4 (with a Spearman correlation of −0.93 between symptom means and standard deviations; for Data Sets 1 through 3, the correlations were −0.63, −0.41, and −0.27, respectively). There were considerable similarities across data sets in their mean symptom profiles (see Table 2): Spearman correlations between the symptom profiles ranged from 0.43 (Data Sets 2 and 3) to 0.82 (Data Sets 1 and 2), with a mean correlation of 0.60 (a plot of the symptom means and variances is available in the Supplemental Material).

Nearly all symptoms had a mean of at least 2 on a scale from 1 to 4. On average, across all four data sets, *Amnes* (7) showed the lowest mean of 2.12, and *Sleep* (12) the highest mean of 3.19. The lowest individual symptom mean was *Flash* (3) with 1.76 in Data Set 3, and the highest was *Sleep* (12) with 3.05 in Data Set 4. Table 2 lists all symptoms and short codes, as well as means and standard deviations.

Data Set 1

Data Set 2
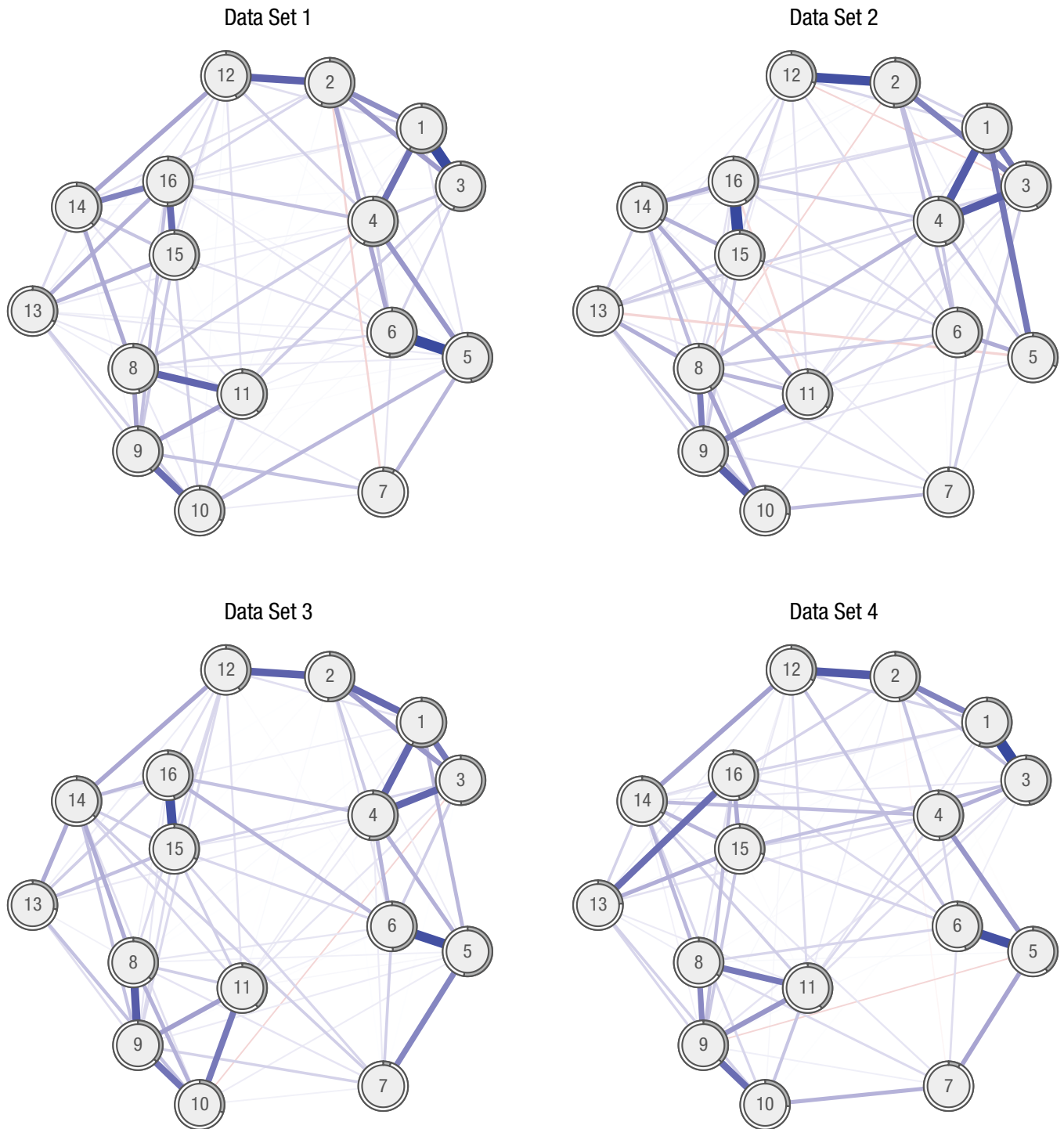
Data Set 3

Data Set 4

Fig. 1. Regularized partial correlation networks across four clinical data sets of traumatized patients receiving treatment. Edge thickness represents the degree of association, blue (solid) edges indicate positive relations, and red (dashed) edges indicate negative relationships. The gray area in the rings around the nodes depicts predictability (the variance of a given node explained by all its neighbors). 1 = intrusions; 2 = nightmares; 3 = flashbacks; 4 = physio-/psychological reactivity; 5 = avoidance of thoughts; 6 = avoidance of situations; 7 = amnesia; 8 = disinterest in activities; 9 = feeling detached; 10 = emotional numbing; 11 = foreshortened future; 12 = sleep problems; 13 = irritability; 14 = concentration problems; 15 = hypervigilance; 16 = startle response.
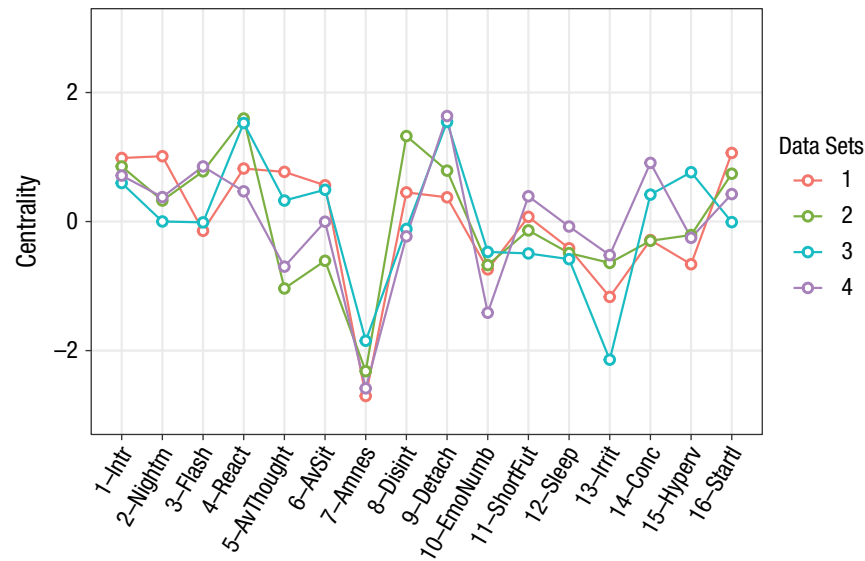
Fig. 2. Standardized node strength centrality of the 16 posttraumatic stress disorder symptoms across four clinical data sets of traumatized patients receiving treatment. See Table 2 for full symptom names.

Network estimation

The four jointly estimated networks are visualized in Figure 1. The four networks featured many consistent edges such as the strong connection between *Nightm* (2)—*Sleep* (12) and the moderate connection between *Detach* (9)—*EmoNumb* (10); in all networks, *Amnes* (7) was weakly interconnected. There were also specific edges that differed considerably across networks, such as *Intr* (1)—*React* (4), which was very weak in Network 4, moderately strong in Networks 1 and 3, and strong in Network 2, or *Startl* (16)—*Hyperv* (15), which was nearly absent in Network 4, moderately strong in Network 1, and strong in Networks 2 and 3. The only moderately strong negative edge that emerged was in Network 3 between *Irrit* (13)—*AvThought* (5), which is not too implausible: People who are less likely to avoid thoughts about the trauma may be more irritable.

Network inference

Strength centrality is shown in Figure 2; the centrality order was substantially related across the four networks, with correlations ranging from 0.63 (Networks 2 and 3) to 0.75 (Networks 2 and 4). *Amnes* (7), *EmoNumb* (10), and *Irrit* (13) had consistently low centrality estimates (all standardized centrality estimates considerably below 0), whereas *Intr* (1), *Detach* (9), and *React* (4) emerged as consistently central symptoms.

Average predictability in the four networks was similar, with mean predictability for the 16 symptoms

ranging from 35% (Data Set 2) to 43% (Data Set 1). This means that, on average, 38.8% of the variance of each node across the data sets was explained by its neighbors. This is somewhat lower than the two subclinical PTSD data sets reported by Haslbeck and Fried (2017). As expected, strength was strongly related to predictability, with correlations of 0.92, 0.80, 0.62, and 0.74 for Networks 1 through 4.

Network stability

Stability analyses indicated that all four networks were accurately estimated, with small to moderate confidence intervals around the edge weights. The correlation-stability coefficient for strength centrality was 0.60, 0.59, 0.75, and 0.52 for Networks 1 through 4, respectively, and thus exceeded the recommended threshold for stable estimation of 0.50 (Epskamp, Borsboom, & Fried, 2017). Details are available in the Supplemental Material.

Network comparison

To obtain a coefficient of similarity for the networks, we correlated the edge weights for each pair of networks with each other. Spearman correlations ranged from 0.62 (Networks 2 and 4) to 0.74 (Networks 1 and 3), indicating strong similarities. We also used the NCT to compare the four networks. In the omnibus tests, all six pairs of networks differed significantly from each other (all $ps < .005$), implying that no pair of networks featured exactly the same 120 edge weights. Because
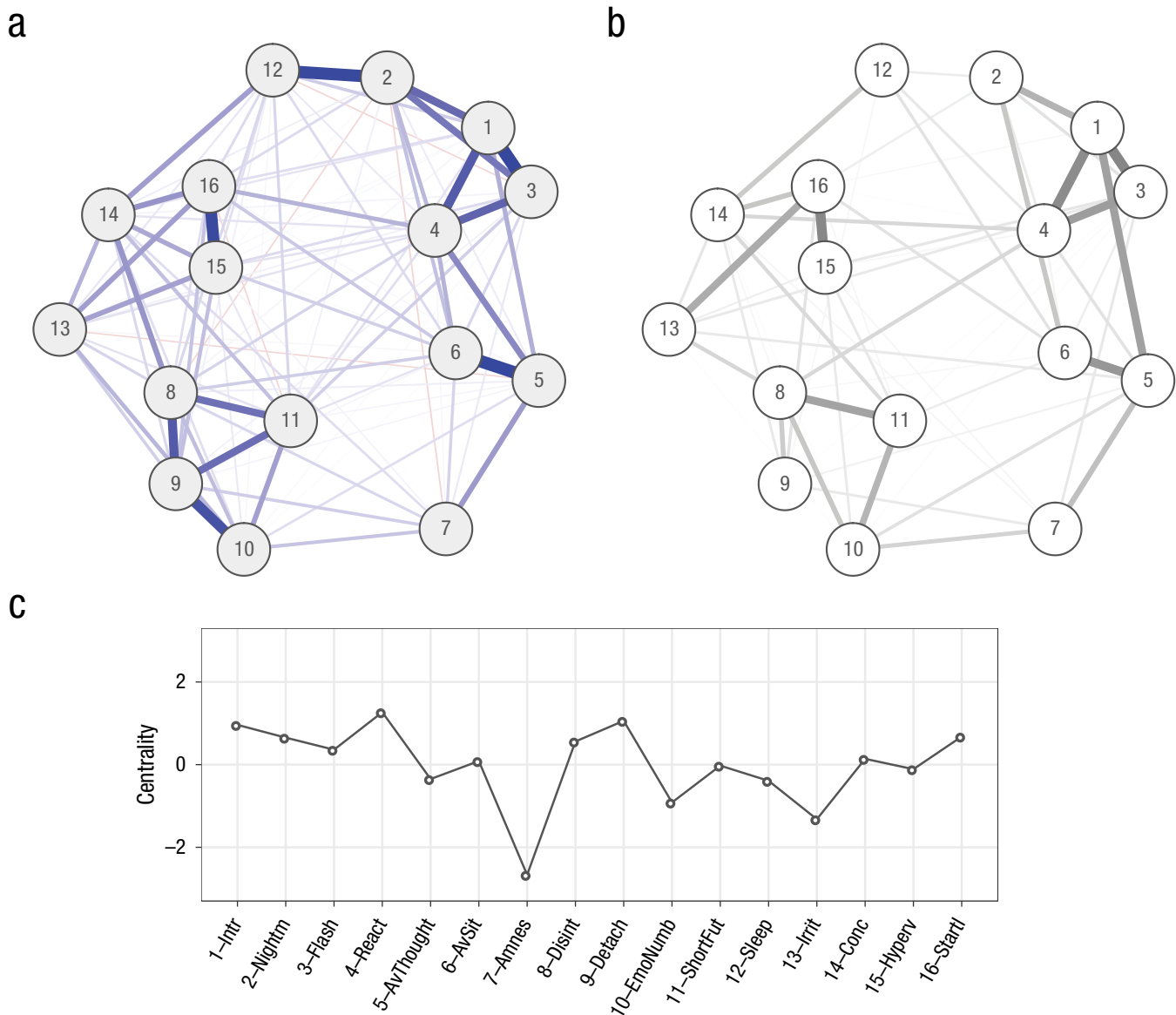
Fig. 3. Network analysis in the combined data set. (a) Cross-sample network (*n* = 2,783) depicts the average of the four individual networks; blue (solid) edges indicate positive relations, and red (dashed) edges indicate negative relationships. In the (b) cross-sample variability network, each edge depicts the standard deviation of this edge across the four networks. (c) Standardized node strength centrality for the cross-sample network is shown. See Figure 2 for node names and Table 2 for full symptom names.

the omnibus test will—given enough power—lead to a significant network difference if only 1 of the 120 edges is different across networks, we also used post hoc tests, comparing all individual edges. Of all 120 edges for each comparison of networks, only 2 (1.7%; comparing Networks 1 vs. 2 and 1 vs. 4) to 8 (6.7%; comparing Networks 3 vs. 4) edges differed significantly across the networks, with a mean of significantly different edges across the six comparisons of 3.1 edges. Details for all significantly different edges are described in the Supplemental Material. Overall, networks were moderately to strongly correlated and only a few significantly different edges emerged, which implies considerable similarities. We also tested whether the global

strength estimates of the four networks (i.e., their connectivity) significantly differed. Global strength values were fairly similar with values of 7.05, 6.59, 7.37, and 6.02 for Networks 1 through 4, respectively. The NCT revealed significant differences for Networks 1 versus 2, 1 versus 4, 2 versus 3, and 3 versus 4.

To get a general sense of the symptom associations and centrality in our large, cross-cultural sample of 2,782 trauma patients, we computed a cross-sample network. Figure 3a depicts this network, Figure 3b depicts the cross-sample variability network, and Figure 3c depicts the strength centrality of the cross-sample network from Figure 3a. The strongest edges emerged between *Intr* (1)—*Flash* (3), *AvoThought* (5)—*AvoSit*

(6), *Nightm* (2)—*Sleep* (12), and *Detach* (9)—*EmoNumb* (10), with edge weights of 0.32, 0.32, 0.31, and 0.26, respectively. The most central symptoms were *React* (4), *Detach* (9), *Intr* (1), and *Disint* (8), with standardized strength estimates of 1.27, 1.06, 0.96, and 0.56; *Amnes* (7), with a value of −2.67, was by far the least central symptom.

In the cross-sample variability network, the most variable edges across the four networks were *Intr* (1)—*Flash* (3), *Hyperv* (15)—*Startl* (16), and *Intr* (1)—*React* (4), with standard deviations of 0.15, 0.15, and 0.14, respectively. For the remaining edges, standard deviations were small to negligible and, like all model parameters in this article, are available in the Supplemental Material.

Discussion

This article is the first empirical investigation of similarities of network structures across four clinical data sets, addressing the considerable concern of replicability in the recent network literature (Borsboom et al., 2017; Epskamp, Borsboom, & Fried, 2017; Forbes et al., 2017a; Fried & Cramer, 2017). Specifically, we estimated networks jointly in four trauma populations that differed in terms of cultural background, trauma type, and severity level. The analyzed samples were larger and more severely traumatized than those investigated in most prior PTSD network studies. Our results can be summarized as follows.

First, whereas data sets differed in overall PTSD severity, the patterns of symptom endorsement were correlated across the four samples; this is interesting given the considerable differences across data sets, especially because different trauma types have been shown to vary in their symptom profiles (Kelley et al., 2009). Sleep problems emerged as the overall most severe symptom, followed by concentration problems and intrusions; amnesia had the lowest severity. Second, whereas the structures of the four networks were not statistically identical (i.e., not all edges were exactly the same), the networks showed moderate to high intercorrelations, as did strength centrality coefficients. Third, we highlighted the most pronounced differences among networks by estimating a variability network: The associations between intrusions and flashbacks, intrusions and physiological/psychological reactivity, and being startled and hypervigilance differed considerably across the four samples, whereas other edges were similar or identical across networks.

In the next sections, we discuss our results in more detail, highlight strengths and limitations of the study, and conclude by outlining future directions for network replicability studies.

Severity and centrality of PTSD symptomatology

If we assume that a symptom is central because it shares a number of outgoing, causal connections with other symptoms—which we can estimate only as undirected edges in cross-sectional analysis—this implies that central symptoms may be especially relevant for treatment. In the current study, psychological reactivity, intrusive traumatic memories, detachment, and disinterest in activities were among the most central symptoms. Whereas most manuals for trauma-focused treatments such as cognitive-behavioral and exposure-based therapies focus on intrusions or reducing anxiety associated with traumatic memories, they do not explicitly focus on disinterest in activities. These treatments are generally effective (Watts et al., 2013), but more than one-third of patients show little to no improvement (Bradley, Greene, Russ, Dutra, & Westen, 2005; Steenkamp, Litz, Hoge, & Marmar, 2015). Future research should therefore aim to elucidate whether further improvements can be reached when treatments also target other central symptoms such as loss of interest by focusing on activation via reinforcement of activities, as is commonly part of behavioral activation in depression treatment. Experimental studies that intervene directly on such central symptoms are needed to test whether this would indeed affect other symptoms as well (Fried & Cramer, 2017).

The high centrality of detachment and disinterest is also noteworthy from another perspective: Contrasting the fifth edition of the *Diagnostic and Statistical Manual of Mental Disorders* (*DSM–5*; American Psychiatric Association, 2013), the ICD-11 removed them as criteria for PTSD. Whereas detachment is now listed as a symptom for complex PTSD, disinterest was removed altogether. First estimates have shown that the ICD-11 criteria lead to substantial reductions in PTSD prevalence compared with rates obtained via *DSM–5* or ICD-10 symptom lists (Wisco et al., 2016). Given the relevance of detachment and disinterest across different trauma samples in our study, we argue that it is safer to err on the side of caution—and thus include these symptoms in research surveys and statistical models—rather than not assessing them in research contexts and missing out on potentially relevant data. This also pertains to symptoms of other disorders relevant in the context of PTSD. The most central symptoms in our study are a mix of "classic" PTSD symptoms (e.g., reactivity and intrusions) and symptoms that are related to depressive disorders (e.g., disinterest and concentration problems), which is especially relevant given the high comorbidity rates between PTSD and major depression (Flory & Yehuda, 2015) and the association of PTSD symptomatology with general negativity and

neuroticism (Engelhard, van den Hout, & Lommen, 2009). The network perspective offers a powerful framework to understand such comorbid conditions by putting the focus on bridge symptoms between disorders (Cramer et al., 2010; Fried et al., 2017). Future studies should aim to unravel the causal associations among such symptoms that cut across diagnostic boundaries. Network theory would predict that patients who develop bridge symptoms may be especially vulnerable to developing comorbid conditions. This means that PTSD patients developing symptoms that are also criteria for major depression, such as sleep or concentration problems, may require special monitoring and offers novel opportunities for prevention research.

Amnesia emerged as a symptom with consistently low severity and centrality across our data sets and networks. Given that centrality reflects the degree of association with other items, we would expect that low-centrality items are those that do not show high factor loadings. This is indeed the case for amnesia, which usually stands out in factor models as a "problematic" item because it does not fit well into the latent structure (Armour, Tsai, et al., 2015; Forbes et al., 2015). From a purely data-driven perspective, in which the idea is to define a syndrome as a list of symptoms that commonly co-occur, amnesia is thus the symptom that fits PTSD the least because—as we have shown here—it occurs less often than other symptoms and also shows weaker correlations with other symptoms. Although a detailed discussion of the symptom is beyond the scope of this article, amnesia is widely acknowledged as one of the most problematic PTSD *DSM* items (McNally, 2009; Rubin, Berntsen, & Bohni, 2008).

Are central symptoms viable intervention targets?

It is important to highlight that centrality does not automatically translate to clinical relevance and that highly central symptoms are not automatically viable intervention targets. Suppose a symptom is central because it is the causal endpoint for many pathways in the data: Intervening on such a product of a causal chain would not lead to any changes in the system. Another possibility is that undirected edges imply feedback loops (i.e., A—B comes from A↔B), in which case a highly central symptom such as insomnia would feature many of these loops. This would make it an intervention target that would have a strong effect on the network if it succeeded—but an intervention with a low success probability, because feedback loops that lead back into insomnia would turn the symptom "on" again after we switch it "off" in therapy. A third example is that a

symptom with the lowest centrality, unconnected to most other symptoms, might still be one of the most important clinical features. No clinician would disregard suicidal ideation or paranoid delusions as unimportant just because they have low centrality values in a network. Another possibility is that a symptom is indeed highly central and causally affects many other nodes in the network but might be very difficult to target in interventions. As discussed by Robinaugh, Millner, and McNally (2016), "Nodes may vary in the extent to which they are amenable to change" (p. 755). Finally, a point we discuss in more detail in the Strengths and Limitations section, centrality can be biased if the shared variance between two nodes derives not from an interaction but from measuring the same latent variable.

In sum, centrality is a metric that needs to be interpreted with great care and in the context of what we know about the sample, the network characteristics, and its elements. If we had to put our money on selecting a clinical feature as an intervention target in the absence of all other clinical information, however, choosing the most central node might be a viable heuristic.

Relation to prior PTSD articles

How do our findings line up with prior PTSD network articles? This is not an easy question to answer for several reasons. There is a considerable number of published articles that have used different symptom sets as a basis for network estimation, including *DSM–IV* symptoms (e.g., McNally et al., 2015), *DSM–5* symptoms (e.g., Afzali, Sunderland, Batterham, et al., 2017; Armour et al., 2016), and other scales such as the 10-item Trauma Screening Questionnaire (Sullivan et al., 2016). These symptom lists differ considerably from each other in length and content, making the integration of our and previous findings challenging. Furthermore, we are aware of only two PTSD articles that made data publicly available (Armour et al., 2016; McNally et al., 2015), and a few articles made the adjacency matrices of their network models available, which makes statistical comparisons of the networks we obtained in our analysis with networks estimated in the prior literature impossible. Differences in sample size may also explain differences in network structures, because regularized partial correlation networks apply regularization procedures that act proportionately to power. When sample size goes to infinity, regularized and unregularized estimation procedures will result in very similar network structures, because even very small edges will be estimated reliably (Epskamp & Fried, 2017; Epskamp, Kruis, & Marsman, 2017). In small samples, however, regularization will set even

moderately large edge weights to zero, resulting in much sparser networks; this further complicates comparisons of network results across articles.

However, we identified one sample that is somewhat similar to our data sets: The population of 362 survivors of the Wenchuan earthquake in China from the article by McNally et al. (2015). The data set is smaller than most of our data sets and covers a different cultural background, and participants did not seek treatment (average symptom severity across our data: 2.76; in McNally data: 1.71 after rescaling 1–5 to 1–4 range). Nevertheless, the authors also used the *DSM–IV* symptom criteria and estimated a regularized partial correlation network in ordinal data. We prepared their data set in the same way we prepared our data (16 instead of 17 symptoms), estimated a GGM, and compared it to our cross-sample network (see Supplemental Material for our code and the data by McNally et al., 2015). The correlation coefficient between the two network structures was 0.51, and the correlation of centrality estimates was 0.55; networks were comparable in terms of overall connectivity (McNally: 7.47; our cross-sample network: 7.15). Although the similarity of network structures is still considerable, given the pronounced differences of data sets, it is substantially lower than the similarity among the four networks we present here.

In general, follow-up work is required to explore differences in network structures and centrality estimates in different PTSD samples, and we hypothesize that differences between our findings and those of McNally et al. (2015) could be attributable to differences in sample size, level of clinical severity, and cultural background.

Strengths and limitations

The particular strengths of the study are its clinical, multisite, and transcultural nature and that we cover a broad spectrum of trauma patients in terms of clinical severity and trauma types. Symptoms were assessed recently, limiting recall bias. We extended the joint network estimation procedure FGL and use it for the first time to estimate four networks jointly. Furthermore, we make available all code and data necessary to fully reproduce our analyses. Most important, this is the very first study to investigate the empirical replicability of PTSD networks across data sets, and the first study ever to investigate network replicability across four data sets.

At the same time, we must acknowledge a number of limitations. We would have preferred to compare data sets on more variables, such as impairment of functioning, or the specific cultures that patients come from (e.g., do PTSD networks differ among refugees from the Middle East versus East Africa?). Unfortunately,

the advantage of pooling data is a disadvantage in this case, because different data sets used different measures or did not assess ethnicity or country of origin with the same level of specificity, precluding us from more detailed comparison. This, to a smaller degree, pertains also to the PTSD scales used: Symptoms were assessed via the HTQ, PSS-SR, and PCL-C that differ in several aspects such as item range (1–4 vs. 1–5), number of items (16 vs. 17), and last-week versus last-month symptom assessment. Note also that assessment took place in Denmark and the Netherlands, and different languages were used when assessing symptomatology. Despite the differences in symptom assessment, the network structures and item mean levels were moderately to highly consistent across data sets.

Comorbidity rates are also among variables we would have preferred to study in more detail, given the considerable prevalence of comorbid disorder in PTSD populations (Kessler, Sonnega, Bromet, Hughes, & Nelson, 1995). Because of the clinical nature of the data sets and their focus on treatment (and not research), not all data sets assessed comorbid disorders, and we were unable to compare comorbidity rates across data sets that may explain differences of networks. For instance, Data Set 4/Network 4, which stands out somewhat from the others, was estimated in a population of refugees with 30% prevalence of persistent psychotic symptoms. Although psychotic symptoms are not uncommon in individuals with PTSD, they might constitute a special PTSD subtype (Braakman, Kortmann, & van den Brink, 2009). Unfortunately, the etiology of PTSD with psychotic symptoms is still poorly understood, and some lines of inquiry indicate that comorbid depression with psychotic symptoms might be responsible for this co-occurrence (Gaudiano & Zimmerman, 2010). Because it is unlikely from a network perspective that symptoms of a given disorder trigger only symptoms of this disorder—especially for diagnoses with high comorbidities (Fried & Cramer, 2017)—future investigations should aim to include a broad range of symptoms in their models. For PTSD, an important step would be to focus on depression and anxiety symptoms as well as psychotic symptoms in severe psychiatric populations.

A final challenge is that specific psychopathology symptoms in networks may measure the same underlying variable. As discussed in detail elsewhere (Fried & Cramer, 2017), if rating scales assess the same symptom with multiple questions, it is questionable whether all should be included in a network analysis because edges are unlikely putative causal pathways. For the 16 PTSD items in our study, this seems potentially relevant for the strong edge between nightmares and sleep problems that could be argued to measure similar content. On the other hand, there is evidence that nightmares

and sleep problems differ from each other in important aspects, which is why we decided to retain both in the analyses. For instance, predeployment nightmares in soldiers predict PTSD symptoms at 6 months postdeployment, whereas predeployment insomnia complaints do not (Van Liempt, Van Zuiden, Westenberg, Super, & Vermetten, 2013), and nightmares more strongly predict future suicides than other sleep problems such as hypersomnia, difficulties initiating sleep, difficulties maintaining sleep, and early morning awakening (Sjöström, Waern, & Hetta, 2007).

Conclusion

Network models have been used as alternate conceptualizations of symptom co-occurrence, contrasting the idea that all symptoms stem from one common cause. Especially for PTSD, however, we need to address the elephant in the room: Trauma can clearly be understood as a common cause for PTSD symptoms. Then again, many causal pathways between symptoms are also plausible. In a recent article, *hybrid models* have been proposed: A common cause is responsible for the onset of PTSD (moderated and mediated by vulnerability and protective factors), whereas the maintenance of the disorder is governed by a network of symptom associations (Fried & Cramer, 2017). This changes the relationship of common cause and network conceptualizations from competing to complementary and offers crucial research opportunities for future work on statistical hybrid models (see Epskamp, Rhemtulla, & Borsboom, 2016).

Cross-sample investigations such as those in this article require considerably more effort to conduct than studies in one data set, which explains why researchers in the clinical network modeling literature have largely refrained from doing so—a practice that poses challenges to the generalizability and replicability of findings (Epskamp, Borsboom, & Fried, 2017; Forbes et al., 2017a; Fried & Cramer, 2017; Tackett et al., 2017). Whereas network structures generalize fairly well across four heterogeneous clinical samples in this article, the open questions are how well PTSD networks generalize to other clinical samples or to community samples and how well networks of other disorders replicate.

When we started the investigation that resulted in this article, no articles were available on cross-sample network replicability. During the revision of this article, two manuscripts were accepted for publication that aimed to address related questions. First, Forbes et al. (2017a) investigated whether different network models estimated on depression and anxiety symptoms replicate across two large community data sets. Unfortunately, the two data sets contain a large proportion of

missing data as a result of skipped questions, which the authors imputed with zeros, a procedure that biases the relationships among variables in the same way in both data sets. This complicates the question of replicability considerably because similarities in network models are driven by similarities of the two correlation matrices, which in turn are strongly influenced by the same skip structure. Additionally, Forbes and colleagues did not always use models appropriate for the data (e.g., they fit relative importance networks based on linear regressions to binary data), did not use state-of-the-art methodology to compare models such as the NCT, and did not estimate all network models correctly (they deleted strong edges from the relative importance networks). For a critical discussion and detailed reanalysis of the article, we refer the reader to the commentary of Borsboom et al. (2017). Second, Verschuere et al. (2017) estimated network models on the basis of psychopathy items in three large clinical offender/forensic samples. They did not, however, formally test the similarity or difference of the network structures and instead focused on whether results of centrality analyses were consistent across the data sets. This article thus stands out from these two prior studies in four aspects: (a) We tested replicability across four data sets; (b) we investigated PTSD network replicability; (c) we used formal psychometric tests to investigate whether network structures differ from each other statistically; and (d) we used a novel estimation framework, the FGL, that is well suited for estimating networks across multiple data sets.

The question of replicability is a challenge not limited to network models and is equally relevant for factor models where researchers commonly explore the factor structure of a given mental disorder such as PTSD or depression using only one data set (for notable exceptions, see Cole et al., 2011; Krueger, Chentsova-Dutton, Markon, Goldberg, & Ormel, 2003; Waszczuk, Kotov, Ruggero, Gamez, & Watson, 2017). Reviews have shown that these data-driven results for specific disorders often do not generalize, regarding both the number and nature of the extracted factors (e.g., PTSD; Armour, Műllerová, et al., 2015; and depression; Gullion & Rush, 1998; Shafer, 2006; van Loo, de Jonge, Romeijn, Kessler, & Schoevers, 2012), and recent articles have called for more replication work especially for such disorder-specific factor models (Waszczuk et al., 2017). Given that both network and factor models in ordinal and continuous data are estimated on the same correlation matrix, and given that network and factor models are mathematically equivalent under a set of conditions (Epskamp, Maris, Waldorp, & Borsboom, 2016; Kruis & Maris, 2016), generalizability problems for one type of model imply generalizability problems for the other

(Borsboom et al., 2017). If the correlation matrix of items differs considerably across two data sets, both factor and network models will pick up on these differences.

We therefore conclude that investing time in more thoroughly conducted cross-sample studies for both network and factor models is warranted in order to facilitate insights about replicability and generalizability. We hope that this article will encourage more researchers to do so and that sharing the correlation matrices of the four clinical data sets will enable further replicability research on these data.

ORCID ID

Eiko I. Fried (ID) https://orcid.org/0000-0001-7469-594X

Declaration of Conflicting Interests

The author(s) declared that there were no conflicts of interest with respect to the authorship or the publication of this article.

Supplemental Material

Additional supporting information may be found at http://journals.sagepub.com/doi/suppl/10.1177/2167702617745092.

Open Practices

All materials have been made publicly available via the Open Science Framework and can be accessed at https://osf.io/9yvxd/. The complete Open Practices Disclosure for this article can be found at http://journals.sagepub.com/doi/suppl10.1177/2167702617745092. This article has received the badge for Open Materials. More information about the Open Practices badges can be found at https://www.psychologicalscience.org/publications/badges.

Notes

1. Network models and factor models are mathematically equivalent under a set of assumptions (Epskamp, Maris, Waldorp, & Borsboom, 2016; Kruis & Maris, 2016), and differences across data sets for one type of model imply differences for the other.
2. Because different sample sizes can lead to loss of power when comparing two networks, we estimated network comparisons also in a different way. For each network comparison, we subsampled the larger data set down to the same size of the smaller data set 5 times each and repeated the *NetworkComparisonTest* procedure as described above. The results were nearly identical, and we thus report the conceptually simpler analysis with unequal samples in the article and the sensitivity analysis in the Supplemental Material.
3. A Bayes factor (BF) of 10 indicates that the data are 10 times more likely under H1 than under H0; a BF of 0.2 indicates that data are 5 times more likely under H0 than under H1. A BF > 100 can be considered very strong evidence for H1 relative to H0, which in our case are mean differences (see Berger, 2006).

References

Afzali, M. H., Sunderland, M., Batterham, P. J., Carragher, N., Calear, A., & Slade, T. (2017). Network approach to the symptom-level association between alcohol use disorder and posttraumatic stress disorder. *Social Psychiatry and Psychiatric Epidemiology*, *52*(3), 329–339. doi:10.1007/s00127-016-1331-3

Afzali, M. H., Sunderland, M., Teesson, M., Carragher, N., Mills, K., & Slade, T. (2017). A network approach to the comorbidity between posttraumatic stress disorder and major depressive disorder: The role of overlapping symptoms. *Journal of Affective Disorders*, *208*, 490–496. doi:10.1016/j.jad.2016.10.037

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Washington, DC: Author.

Armour, C., Fried, E. I., Deserno, M. K., Tsai, J., Pietrzak, R. H., & Southwick, S. M. (2016). A network analysis of *DSM-5* posttraumatic stress disorder symptoms and correlates in U.S. military veterans. *Journal of Anxiety Disorders*, *45*, 49–59. doi:10.1016/j.janxdis.2016.11.008

Armour, C., Műllerová, J., & Elhai, J. D. (2015). 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. doi:10.1016/j.cpr.2015.12.003

Armour, C., Tsai, J., Durham, T. A., Charak, R., Biehn, T. L., Elhai, J. D., & Pietrzak, R. H. (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. doi:10.1016/j.jpsychires.2014.10.012

Beard, C., Millner, A. J., Forgeard, M. J. C., Fried, E. I., Hsu, K. J., Treadway, M., . . . Björgvinsson, T. (2016). Network analysis of depression and anxiety symptom relations in a psychiatric sample. *Psychological Medicine*, *46*(16), 3359–3369. doi:10.1017/S0033291716002300

Berger, J. O. (2006). Bayes factors. In S. Kotz, N. Balakrishnan, C. Read, B. Vidakovic, & N. L. Johnson (Eds.), *Encyclopedia of statistical sciences* (2nd ed., Vol. 1, pp. 378–386). Hoboken, NJ: Wiley.

Birkeland, M. S., & Heir, T. (2017). Making connections: Exploring the centrality of posttraumatic stress symptoms and covariates after a terrorist attack. *European Journal of Psychotraumatology*, *8*(Suppl. 3), 1333387. doi:10.1080/2000 8198.2017.1333387

Borsboom, D. (2017). A network theory of mental disorders. *World Psychiatry*, *16*, 5–13.

Borsboom, D., & Cramer, A. O. J. (2013). Network analysis: An integrative approach to the structure of psychopathology. *Annual Review of Clinical Psychology*, *9*, 91–121. doi:10.1146/annurevclinpsy-050212-185608

Borsboom, D., Fried, E. I., Epskamp, S., Waldorp, L. J., van Borkulo, C. D., van der Maas, H. L. J., & Cramer, A. O. J. (2017). False alarm? A comprehensive reanalysis of "Evidence that psychopathology symptom networks have limited replicability" by Forbes, Wright, Markon, and Krueger (2017). *Journal of Abnormal Psychology*, *126*(7), 989–999.

Braakman, M. H., Kortmann, F. A. M., & van den Brink, W. (2009). Validity of "post-traumatic stress disorder with secondary psychotic features": A review of the evidence. *Acta Psychiatrica Scandinavica*, *119*(1), 15–24. doi:10.1111/j.1600-0447.2008.01252.x

Bradley, R., Greene, J., Russ, E., Dutra, L., & Westen, D. (2005). A multidimensional meta-analysis of psychotherapy for PTSD. *The American Journal of Psychiatry*, *162*(2), 214–227. doi:10.1176/appi.ajp.162.2.214

Bringmann, L. F., Vissers, N., Wichers, M., Geschwind, N., Kuppens, P., Peeters, F., . . . Tuerlinckx, F. (2013). A network approach to psychopathology: New insights into clinical longitudinal data. *PLOS ONE*, *8*(4), Article e60188. doi:10.1371/journal.pone.0060188

Bryant, R. A., Creamer, M., O'Donnell, M., Forbes, D., McFarlane, A. C., Silove, D., & Hadzi-Pavlovic, D. (2017). Acute and chronic posttraumatic stress symptoms in the emergence of posttraumatic stress disorder. *JAMA Psychiatry*, *74*(2), 135–142. doi:10.1001/jamapsychiatry .2016.3470

Carlsson, J., Sonne, C., & Silove, D. (2014). From pioneers to scientists: Challenges in establishing evidence-gathering models in torture and trauma mental health services for refugees. *The Journal of Nervous and Mental Disease*, *202*(9), 630–637. doi:10.1097/NMD.0000000000000175

Cole, D. A., Cai, L., Martin, N. C., Findling, R. L., Youngstrom, E. A., Garber, J., . . . Forehand, R. (2011). Structure and measurement of depression in youths: Applying item response theory to clinical data. *Psychological Assessment*, *23*(4), 819–833. doi:10.1037/a0023518

Costantini, G., & Epskamp, S. (2017). EstimateGroupNetwork: Perform the joint graphical lasso and selects tuning parameters. R package (Version 0.1.2) [Computer software]. Retrieved from https://cran.r-project.org/web/packages/ EstimateGroupNetwork/index.html

Costantini, G., & Perugini, M. (2017). Network analysis for psychological situations. In D. C. Funder, J. F. Rauthmann, & R. A. Sherman (Eds.), *The Oxford handbook of psychological situations*. New York, NY: Oxford University Press. doi:10.1093/oxfordhb/ 9780190263348.013.16

Costantini, G., Richetin, J., Preti, E., Casini, E., Epskamp, S., & Perugini, M. (2017). Stability and variability of personality networks. A tutorial on recent developments in network psychometrics. *Personality and Individual Differences*. Advance online publication. doi:10.1016/j.paid.2017.06.011

Cramer, A. O. J., Waldorp, L. J., van der Maas, H. L. J., & Borsboom, D. (2010). Comorbidity: A network perspective. *The Behavioral and Brain Sciences*, *33*(2–3), 137–150. doi:10.1017/S0140525X09991567

Danaher, P., Wang, P., & Witten, D. M. (2014). The joint graphical lasso for inverse covariance estimation across multiple classes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *76*(2), 373–397. doi:10.1111/rssb.12033

Engelhard, I. M., van den Hout, M. A., & Lommen, M. J. J. (2009). Individuals high in neuroticism are not more reactive to adverse events. *Personality and Individual Differences*, *47*(7), 697–700. doi:10.1016/j.paid.2009.05.031

Epskamp, S., Borsboom, D., & Fried, E. I. (2017). Estimating psychological networks and their accuracy: A tutorial paper. *Behavior Research Methods*. Advance online publication. doi:10.3758/s13428-017-0862-1

Epskamp, S., Cramer, A. O. J., Waldorp, L. J., Schmittmann, V. D., & Borsboom, D. (2012). qgraph: Network visualizations of relationships in psychometric data. *Journal of Statistical Software*, *48*(4), 1–18.

Epskamp, S., & Fried, E. I. (2017). A tutorial on regularized partial correlation networks. *Psychological Methods*. doi:arxiv.org/abs/1607.01367

Epskamp, S., Kruis, J., & Marsman, M. (2017). Estimating psychopathological networks: Be careful what you wish for. *PLOS ONE*, *12*(6), Article e0179891. doi:10.1371/journal .pone.0179891

Epskamp, S., Maris, G., Waldorp, L. J., & Borsboom, D. (2016). Network psychometrics. In P. Irwing, D. Hughes, & T. Booth (Eds.), *Handbook of psychometrics*. New York, NY: Wiley. Retrieved from https://arxiv.org/abs/1609.02818

Epskamp, S., Rhemtulla, M., & Borsboom, D. (2017). Generalized network psychometrics: Combining network and latent variable models. *Psychometrika*, *82*, 904–927. doi:10.1007/s11336-017-9557-x

Flory, J. D., & Yehuda, R. (2015). Comorbidity between posttraumatic stress disorder and major depressive disorder: Alternative explanations and treatment considerations. *Dialogues in Clinical Neuroscience*, *17*(2), 141–150.

Foa, E. B., Cashman, L., Jaycox, L., & Perry, K. (1997). The validation of a self-report measure of posttraumatic stress

disorder: The Posttraumatic Diagnostic Scale. *Psychological Assessment*, *9*(4), 445–451. doi:10.1037/1040-3590.9.4.445

Forbes, D., Lockwood, E., Elhai, J. D., Creamer, M., Bryant, R., McFarlane, A., . . . O'Donnell, M. (2015). An evaluation of the *DSM-5* factor structure for posttraumatic stress disorder in survivors of traumatic injury. *Journal of Anxiety Disorders*, *29*, 43–51. doi:10.1016/j.janxdis.2014.11.004

Forbes, M. K., Wright, A. G. C., Markon, K. E., & Krueger, R. F. (2017a). Evidence that psychopathology symptom networks have limited replicability. *Journal of Abnormal Psychology*, *126*(7), 969–988.

Forbes, M. K., Wright, A. G. C., Markon, E. K., & Krueger, R. F. (2017b). Further evidence that psychopathology networks have limited replicability and utility: Response to Borsboom et al. (2017) and Steinley et al. (2017). *Journal of Abnormal Psychology*, *126*(7), 1011–1016.

Forbush, K. T., Siew, C. S. Q., & Vitevitch, M. S. (2016). Application of network analysis to identify interactive systems of eating disorder psychopathology. *Psychological Medicine*, *46*(12), 2667–2677. doi:10.1017/S003329171600012X

Frewen, P. A., Schmittmann, V. D., Bringmann, L. F., & Borsboom, D. (2013). Perceived causal relations between anxiety, posttraumatic stress and depression: Extension to moderation, mediation, and network analysis. *European Journal of Psychotraumatology*, *4*, 20656. doi:10.3402/ejpt.v4i0.20656

Fried, E. I., & Cramer, A. O. J. (2017). Moving forward: Challenges and directions for psychopathological network theory and methodology. *Perspectives on Psychological Science*, *12*, 999–1020. doi:10.1177/1745691617705892

Fried, E. I., Epskamp, S., Nesse, R. M., Tuerlinckx, F., & Borsboom, D. (2016). What are "good" depression symptoms? Comparing the centrality of *DSM* and non-*DSM* symptoms of depression in a network analysis. *Journal of Affective Disorders*, *189*, 314–320. doi:10.1016/j.jad.2015.09.005

Fried, E. I., van Borkulo, C. D., Cramer, A. O. J., Boschloo, L., Schoevers, R. A., & Borsboom, D. (2017). Mental disorders as networks of problems: A review of recent insights. *Social Psychiatry and Psychiatric Epidemiology*, *52*(1), 1–10. doi:10.1007/s00127-016-1319-z

Friedman, J., Hastie, T., & Tibshirani, R. (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, *9*, 432–441. doi:10.1093/biostatistics/kxm045

Gaudiano, B. A., & Zimmerman, M. (2010). Evaluation of evidence for the psychotic subtyping of post-traumatic stress disorder. *The British Journal of Psychiatry*, *197*(4), 326–327. doi:10.1192/bjp.bp.110.078824

Gullion, C. M., & Rush, A. J. (1998). Toward a generalizable model of symptoms in major depressive disorder. *Biological Psychiatry*, *44*(10), 959–972. doi:10.1016/S0006-3223(98)00235-2

Haslbeck, J. M. B., & Fried, E. I. (2017). How predictable are symptoms in psychopathological networks? A reanalysis of 18 published datasets. *Psychological Medicine*, *47*(16), 2767–2776. doi:10.1017/S0033291717001258

Hayes, A. M., Yasinski, C., Ben Barnes, J., & Bockting, C. L. H. (2015). Network destabilization and transition in depression: New methods for studying the dynamics of therapeutic change. *Clinical Psychology Review*, *41*, 27–39. doi:10.1016/j.cpr.2015.06.007

Isvoranu, A.-M., Borsboom, D., van Os, J., & Guloksuz, S. (2016). A network approach to environmental impact in psychotic disorder: Brief theoretical framework. *Schizophrenia Bulletin*, *42*(4), 870–873. doi:10.1093/schbul/sbw049

Karstoft, K. I., Andersen, S. B., Bertelsen, M., & Madsen, T. (2014). Diagnostic accuracy of the Posttraumatic Stress Disorder Checklist–Civilian Version in a representative military sample. *Psychological Assessment*, *26*(1), 321–325. doi:10.1037/a0034889

Kelley, L., Weathers, F., McDevitt-Murphy, M. E., Eakin, D., & Flood, A. (2009). A comparison of PTSD symptom patterns in three types of civilian trauma. *Journal of Traumatic Stress*, *22*(3), 227–235. doi:10.1002/jts

Kessler, R. C., Sonnega, A., Bromet, E., Hughes, M., & Nelson, C. B. (1995). Posttraumatic stress disorder in the National Comorbidity Survey. *Archives of General Psychiatry*, *52*(12), 1048–1060. doi:10.1001/archpsyc.1995.03950240066012

Knefel, M., Tran, U. S., & Lueger-Schuster, B. (2016). The association of posttraumatic stress disorder, complex posttraumatic stress disorder, and borderline personality disorder from a network analytical perspective. *Journal of Anxiety Disorders*, *43*, 70–78. doi:10.1016/j.janxdis.2016.09.002

Krueger, R. F., Chentsova-Dutton, Y. E., Markon, K. E., Goldberg, D., & Ormel, J. (2003). A cross-cultural study of the structure of comorbidity among common psychopathological syndromes in the general health care setting. *Journal of Abnormal Psychology*, *112*(3), 437–447. doi:10.1037/0021-843X.112.3.437

Kruis, J., & Maris, G. (2016). Three representations of the Ising model. *Scientific Reports*, *6*, 34175. doi:10.1038/srep34175

Lauritzen, S. L. (1996). *Graphical models*. Gloucestershire, England: Clarendon Press.

Marsella, A. J., Matthew, J., Friedman, M. J., Gerrity, E., & Scurfield, R. M. (1996). *Ethnocultural aspects of posttraumatic stress disorders: Issues, research and applications*. Washington, DC: American Psychological Association.

McNally, R. J. (2009). Can we fix PTSD in *DSM-V*? *Depression and Anxiety*, *26*(7), 597–600. doi:10.1002/da.20586

McNally, R. J. (2016). Can network analysis transform psychopathology? *Behaviour Research and Therapy*. doi:10.1016/j.brat.2016.06.006

McNally, R. J., Robinaugh, D. J., Wu, G. W. Y., Wang, L., Deserno, M. K., & Borsboom, D. (2015). Mental disorders as causal systems: A network approach to posttraumatic stress disorder. *Clinical Psychological Science*, *3*(3), 836–849.

Mitchell, K. S., Wolf, E. J., Bovin, M. J., Lee, L. O., Green, J. D., Raymond, C., . . . Marx, B. P. (2017). Network models of DSM-5 posttraumatic stress disorder: Implications for ICD-11. *Journal of Abnormal Psychology*, *126*(3), 355–366.

Mollica, R., Caspi-Yavin, Y., Bollini, P., Truong, T., Tor, S., & Lavelle, J. (1992). The Harvard Trauma Questionnaire: Adapting a cross-cultural instrument for measuring torture, trauma and posttraumatic stress disorder in Iraqi refugees. *The Journal of Nervous & Mental Disease*, *180*(2), 111–116.

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science, 349*(6251), aac4716. doi:10.1126/science.aac4716

Rhemtulla, M., Fried, E. I., Aggen, S. H., Tuerlinckx, F., Kendler, K. S., & Borsboom, D. (2016). Network analysis of substance abuse and dependence symptoms. *Drug and Alcohol Dependence, 161*, 230–237. doi:10.1016/j.drugalcdep.2016.02.005

Richetin, J., Preti, E., Costantini, G., & De Panfilis, C. (2017). The centrality of affective instability and identity in borderline personality disorder: Evidence from network analysis. *PLOS ONE, 12*(10), Article e0186695. doi:10.1371/journal.pone.0186695

Robinaugh, D. J., Millner, A. J., & McNally, R. J. (2016). Identifying highly influential nodes in the complicated grief network. *Journal of Abnormal Psychology, 125*(6), 747–757. doi:10.1037/abn0000181

Rubin, D. C., Berntsen, D., & Bohni, M. K. (2008). A memory-based model of posttraumatic stress disorder: Evaluating basic assumptions underlying the PTSD diagnosis. *Psychological Review, 115*(4), 985–1011. doi:10.1037/a0013397

Shafer, A. B. (2006). Meta-analysis of the factor structures of four depression questionnaires: Beck, CES-D, Hamilton, and Zung. *Journal of Clinical Psychology, 62*(1), 123–146. doi:10.1002/jclp

Sjöström, N., Waern, M., & Hetta, J. (2007). Nightmares and sleep disturbances in relation to suicidality in suicide attempters. *Sleep, 30*(1), 91–95.

Spiller, T. R., Schick, M., Schnyder, U., Bryant, R. A., Nickerson, A., & Morina, N. (2017). Symptoms of posttraumatic stress disorder in a clinical sample of refugees: A network analysis. *European Journal of Psychotraumatology, 8*(Suppl. 2), 1318032. doi:10.1080/20008198.2017.1318032

Steenkamp, M. M., Litz, B. T., Hoge, C. W., & Marmar, C. R. (2015). Psychotherapy for military-related PTSD: A review of randomized clinical trials. *Journal of the American Medical Association, 314*(5), 489–500. doi:10.1001/jama.2015.8370

Sullivan, C. P., Smith, A. J., Lewis, M., & Jones, R. T. (2016). Network analysis of PTSD symptoms following mass violence. *Psychological Trauma: Theory, Research, Practice and Policy*. doi:10.1037/tra0000237

Tackett, J. L., Lilienfeld, S. O., Patrick, C. J., Johnson, S. L., Krueger, R. F., Miller, J. D., . . . Shrout, P. E. (2017). It's time to broaden the replicability conversation: Thoughts for and from clinical psychological science. *Perspectives on Psychological Science, 12*(5), 742–756. doi:10.1177/1745691614549257

van Borkulo, C. D., Borsboom, D., Epskamp, S., Blanken, T. F., Boschloo, L., Schoevers, R. A., & Waldorp, L. J. (2014). A new method for constructing networks from binary data. *Scientific Reports, 4*, 5918. doi:10.1038/srep05918

van Borkulo, C. D., Boschloo, L., Kossakowski, J. J., Tio, P., Schoevers, R. A., Borsboom, D., . . . Boschloo, L. (2017, March). *Comparing network structures on three aspects* (Working Paper). doi:10.13140/RG.2.2.29455.38569

Van Liempt, S., Van Zuiden, M., Westenberg, H., Super, A., & Vermetten, E. (2013). Impact of impaired sleep on the development of PTSD symptoms in combat veterans: A prospective longitudinal cohort study. *Depression and Anxiety, 30*(5), 469–474. doi:10.1002/da.22054

van Loo, H. M., de Jonge, P., Romeijn, J.-W., Kessler, R. C., & Schoevers, R. A. (2012). Data-driven subtypes of major depressive disorder: A systematic review. *BMC Medicine, 10*(1), 156. doi:10.1186/1741-7015-10-156

Verschuere, B., van Ghesel Grothe, A. S., Waldorp, L., Watts, A. L., Lilienfeld, S. O., Edens, J. F., . . . Noordhof, A. (2017). What features of psychopathy might be central? A network analysis of the Psychopathy Checklist-Revised (PCL-R) in three large samples. *Journal of Abnormal Psychology*. Advance online publication. doi:10.1037/abn0000315

Waszczuk, M. A., Kotov, R., Ruggero, C., Gamez, W., & Watson, D. (2017). Hierarchical structure of emotional disorders: From individual symptoms to the spectrum. *Journal of Abnormal Psychology, 126*(5), 613–634. doi:10.1037/abn0000264

Watts, B. V., Schnurr, P. P., Mayo, L., Young-Xu, Y., Weeks, W. B., & Friedman, M. J. (2013). Meta-analysis of the efficacy of treatments for posttraumatic stress disorder. *The Journal of Clinical Psychiatry, 74*(6), e541–e550. doi:10.4088/JCP.12r08225

Weathers, F. W., Litz, B. T., Herman, D. S., Huska, J. A., & Keane, T. M. (1993, January). *The PTSD Checklist (PCL): Reliability, validity, and diagnostic utility*. Poster session presented at the annual convention of the International Society for Traumatic Stress Studies, San Antonio, TX. Retrieved from https://www.researchgate.net/publication/291448760_The_PTSD_Checklist_PCL_Reliability_validity_and_diagnostic_utility

Wisco, B. E., Miller, M. W., Wolf, E. J., Kilpatrick, D., Resnick, H. S., Badour, C. L., . . . Friedman, M. J. (2016). The impact of proposed changes to ICD-11 on estimates of PTSD prevalence and comorbidity. *Psychiatry Research, 240*, 226–233. doi:10.1016/j.psychres.2016.04.043