American Journal of Orthopsychiatry
2014, Vol. 84, No. 2, 134–141

DOI: 10.1037/h0099385

The Effect of the Postdisaster Context on the Assessment of Individual Mental Health Scores

Tim R. Wind
HealthNet TPO, Amsterdam, Noord-Holland, the Netherlands and Arq Psychotrauma Expert Group, Diemen, Noord-Holland, the Netherlands

Pooran C. Joshi
University of Delhi

Rolf J. Kleber
Utrecht University and Arq Psychotrauma Expert Group, Diemen, Noord-Holland, the Netherlands

Ivan H. Komproe
Utrecht University and HealthNet TPO, Amsterdam, Noord-Holland, the Netherlands

Many scholars question the immense variation in rates of mental health outcomes across disaster studies. This study explains this variation by putting forward 2 methodological problems that are inherent to the effect of a disaster context on mental health screening scores. The Hopkins Symptom Checklist-25 was administered in a flood-affected group (n = 318) and a nonaffected group (n = 304) in Uttar Pradesh, India. The affected group showed much higher mean scores on subscales of anxiety and depression. However, factor analyses (i.e., confirmatory factor analyses [CFA] and multilevel confirmatory factor analyses [MCFA]; Muthén, 1994) revealed 2 methodological phenomena that account for the differences in scores. First, the outcomes revealed that a large proportion of covariance between observed mental health variables did not refer to the latent concepts of interest (depression and anxiety), but to the context of both groups (disaster affected vs. nonaffected). The shared effect of the disaster on the context explained a large proportion of the covariances between the items and biased outcomes. Second, after dissecting this group variance, the construct validity of the assessments of anxiety and depression was revealed to be poor and unstable across both groups. The subscales of anxiety and depression referred to different concepts in both groups. These 2 methodological problems have not been discussed thus far, but they contribute to the variation in mental health outcomes across disaster studies.

There is enormous variation in rates of mental health outcomes across disaster studies. These rates range from no mental health problems at all (Scott, Knoth, Beltran-Quiones, & Gomez, 2003) up to 90% of the affected population suffering from mental health problems (Leon, 2004). Rodin and van Ommeren (2009) distinguish two lines of explanations for this variation across studies. First and most obvious, the severity of disasters differs. Therefore, the degree to which affected individuals perceive disasters as traumatic varies, as well as related mental health problems. However, variation in individual disaster experiences only accounts for a small proportion of the variance in mental health problems (Yehuda & McFarlane, 1995). Second, scholars advocated that beyond the severity of disasters per se, the vast methodological differences in disaster studies—such as differences in research designs and sample sizes—also explain variation in mental health outcomes across studies (Galea, Maxwell, & Norris, 2008; Kessler & Wittchen, 2008; Rodin & van Ommeren, 2009).

Two methodological problems that have not been discussed in the disaster literature thus far are related to the effect of the disaster environment on screening scores. Disasters typically create material destruction and loss of social capital (Kawachi & Subramanian, 2006; Wind, Fordham, & Komproe, 2011). Mental health outcomes in the wake of disasters are largely defined by this destructive effect of disasters on the context (Kawachi & Subramanian, 2006; Wind et al., 2011; Wind & Komproe, 2012). When

Tim R. Wind, Department of Research and Development, HealthNet TPO, Amsterdam, Noord-Holland, the Netherlands and Department of Research, Arq Psychotrauma Expert Group, Diemen, Noord-Holland, the Netherlands; Pooran C. Joshi, Department of Anthropology, University of Delhi, Delhi, India; Rolf J. Kleber, Department of Clinical & Health Psychology, Utrecht University, Utrecht, the Netherlands and Department of Research, Arq Psychotrauma Expert Group; Ivan H. Komproe, Department of Cultural Anthropology, Utrecht University and Department of Research and Development, HealthNet TPO.

This study was conducted as part of the MICRODIS project funded by the European Commission under the 6th Framework Programme (036877-2). The authors are most grateful to the participants of our study.

Correspondence concerning this article should be addressed to Tim R. Wind, HealthNet TPO, Department of Research and Development, 1074VJ, Amsterdam, The Netherlands. E-mail: twind@healthnettpo.org

scholars use screening instruments and ignore factors that operate at the individual and postdisaster contextual level, this may have unwanted consequences for the interpretation of individual mental health outcomes.

First, the basic idea of screening instruments is that the covariance between observed variables of screening instruments is determined by the latent mental health construct they refer to (Irvine & Caroll, 1980). However, this assumption is questionable when a shared context influences individual observed scores (see, e.g., neighborhood studies; Wind & Komproe, 2012). The consequence could be that the covariance between the items of the questionnaire that should refer to the latent mental health concept of interest may also refer to answering tendencies as a result of the common or shared living experiences in the same eroded context (Dyer, Hanges, & Hall, 2005). When the latter source of covariance is ignored, this proportion of the covariance is unintentionally attributed to the effect of the underlying mental health concept on the answering of items (Muthén, 1994). The result is that the assessment of mental health outcomes can be biased (Kreft & De Leeuw, 1998). This bias may contribute to unwanted variation of outcomes across disaster studies. The second problem is that without dissecting variance that is related to factors operating at different contextual levels, one cannot be certain that a screening instrument has adequate construct validity. Without fulfillment of the requirement of measurement equivalence of the assessment tools, interpretations of differences between scores across groups or settings may lead to erroneous conclusions (Poortinga, 1975); that is, mental health scores differ across groups whereas, in fact, the underlying concepts may differ.

The goal of the case study presented here is to illustrate these two methodological problems by assessing symptoms of anxiety and depression with the Hopkins Symptom Checklist-25 (HSCL-25) screener (Derogatis, Lipman, Rickels, Uhlenhuth, & Covi, 1974; Lipman, Covi, & Shapiro, 1979) among a disaster and a nonaffected group in Northern India.

Method

Sample

The study presented here took place as part of the MICRODIS research project. MICRODIS is a European-Community-funded research project on the effect of natural disasters (e.g., Wind et al., 2011). In scope of this project, a study was conducted in Uttar Pradesh, India, with a research focus on the effect of natural disasters on mental health.

The Bahraich District, in Uttar Pradesh, India, is annually hit by floods, as in July and August, 2008. We compared a disaster-affected group with a nonaffected group in the region in October 2008. The affected region is situated between the river and a dam. The region on the other side of the dam was unaffected and was identified as a nonaffected group. A multistage random sampling procedure was used to first select four *Gram Panchayats* (smallest political units in the region) in the affected and the nonaffected region and then a sample of households. The sampling procedure resulted in a multilevel data structure: households (level 1), Gram Panchayats (level 2), and region (affected vs. nonaffected; level 3). In the affected group, 318 of 380 respondents participated in the study, and in the nonaffected group 304 of 330 respondents participated in this study. The response rates were 84% and 92%, respectively. The demographics of the samples are depicted in Table 1.

Table 1. *Demographics and Mean Scores and Standard Deviations of Anxiety and Depression*

Demographic	Flood-Affected sample (n = 318)	Control sample (n = 304)
Gender (%)		
Female	39	44
Male	61	56
Mean age (*SD*), years	46.03 (15.74)	47.23 (13.92)
Literacy (%)		
Illiterate	64	53
Literate	36	47
Education (%)		
No education	73	66
Primary education	11	16
Secondary education	11	10
Higher secondary education	4	7
Graduate	1	1
Years of education (*SD*)	2.17 (3.70)	2.45 (3.65)
Religion (%)		
Hindu	92	72
Muslim	8	28
Anxiety (*SD*)	2.52 (.63)	1.92 (.67)
Depression (*SD*)	2.48 (.40)	1.89 (.56)

Instrument

Indicators of mental health, anxiety and depression, were measured by the HSCL-25 (Derogatis et al., 1974; Lipman et al., 1979). The HSCL-25 is composed of a 10-item subscale for anxiety and a 15-item subscale for depression. An item concerning sexual interest was preventively omitted because of the taboo associated with talking about sexual issues.

The HSCL-25 was not available in the language spoken in Northern India (Hindi). Nonetheless, the HSCL-25 has been previously used in the same region of Uttar Pradesh, India (Wind, Joshi, Kleber, & Komproe, 2013) and in the vicinity (among Tibetans in India; Crescenzi et al., 2002). To ensure the cross-cultural validity of the instrument, we used the extensive method of translation described by van Ommeren et al. (1999). This method includes (a) translation by indigenous translators; (b) evaluation by a bilingual professional in terms of comprehensibility, acceptability, relevance, and completeness; (c) evaluation by a focus group of local lay people; (d) blind backtranslation (no differences between the original and the translated version were found); and (e) a 2-day training of the surveyors to administer the instrument.

In the affected sample, the Cronbach's αs of anxiety and depression were, respectively, .81 and .69. In the control sample, the Cronbach's αs of anxiety and depression were, respectively, .90 and .89. These psychometric values in both groups concur with results of studies in western (e.g., Winokur, Winokur, Rickels, & Cox, 1984) and nonwestern settings (e.g., Crescenzi et al., 2002). The difference in Cronbach's αs is considered as a first indication of the potential bias we discussed earlier.

Procedures

Students of the University of Delhi and the Lucknow University familiar with the local sociocultural context and dialect administered the survey under the close supervision of the local principal investigator Joshi (author). The students received a 2-day training in the administration of the HSCL-25 as part of the MICRODIS interview. If possible, written informed consent was obtained. In the case of illiteracy, verbal informed consent and thumb impression were attained and recorded by a witness.

The ethical approval for the study was obtained from the ethical committee of the Department of Anthropology, University of Delhi. The study has been performed in accordance with the ethical guidelines of the Declaration of Helsinki (World Medical Assembly Declaration of Helsinki, 1997).

Statistical Analyses

Before the analyses, the individual responses to the items were screened to determine the normality of the dataset. We used the Shapiro-Wilk test for this purpose: A $p < .05$ refers to a significant deviation from a normal distribution. We computed descriptive statistics for demographic variables and indicators of mental health and used student t tests to examine the differences in mean item scores on the HSCL-25 subscales for anxiety and depression between the affected and the control group. All statistics tests were calculated with SPSS 16.0.

Construct Validity

First, we established the construct validity of the factor structure of the HSCL-25 in the total dataset; therefore, we tested a series of factor models for relative fit: (a) a one-factor model on which all items loaded; (b) an orthogonal two-factor model for the items of the subscales of anxiety and depression; and (c) an oblique two-factor model for the items of the subscales of anxiety and depression. On the basis of the results of Step 1, we used the best-fitting factor structure out of the three specified factor models in the subsequent steps. Goodness of fit measures in the confirmatory factor analysis (CFA; and multilevel confirmatory factor analysis [MCFA]) in this study were (a) the χ^2 test, (b) the root mean square error of approximation (RMSEA), and (c) the comparative fit index (CFI). The χ^2 test is a global test that compares a reconstructed variance/covariance matrix (based on the tested model) with the original variance/covariance matrix of the study sample (Jöreskog & Sörbom, 1993). The RMSEA refers to the misfit of the model and should be less than .06 (Brown, 2006; Browne & Cudeck, 1993; Hu & Bentler, 1999). A CFI $>$ .95 indicates good fit of the model with the data matrix and values in the range of .90–.95 may be indicative of acceptable model fit (Bentler, 1990; Brown, 2006).

Further, we performed multisample CFA to evaluate equality of factor structures by testing a series of hypotheses about the robustness of the factor structure across groups (Jöreskog & Sörbom, 1993). Similarity of patterns of factor loadings can be defined on different levels; thus, there are different hypotheses to test the similarity of factorial composition (Bollen, 1991; Jöreskog & Sörbom, 1993). We tested the different hypotheses of factorial invariance by comparing the absolute fit of different factor models (Byrne, Shavelson, & Muthén, 1989). In this study, we distinguished the following hierarchical models: (a) a model in which the pattern of factor structure is equal across samples (Model A); (b) Model A with the additional constraint that the factor loadings are equal across samples (Model B); (c) Model B with the additional constraint that the error variances are equal across samples (Model C); and (d) Model C with the covariance of the factor items equal across samples (Model D). The difference in χ^2 values among (a) Model A and Model B, (b) Model B and Model C, and (c) Model C and Model D was computed.

The degree of dissimilarity between factor structures across samples determines the difference in the χ^2 between both test models (Devins et al., 1988). When the difference in the χ^2 value of models, $\Delta\chi^2$, is not significant, the hypothesis of invariant factor loadings is tenable (Jöreskog & Sörbom, 1993).

Finally, we applied a four-step procedure of MCFA to identify the proportion of the covariance between observed items that refers to a shared context level (Dyer et al., 2005). In the procedure, within-group variance (i.e., variance relevant for the mental health constructs at the individual level) is distinguished from between-group variance (i.e., nested variance across groups; Muthén, 1990, 1994).

Step 1: Nested variance. In the first step, we estimated the proportion of nested variance for the items of the subscales of anxiety and depression. Hereto, we calculated the intraclass correlation coefficient (ICC) on the basis of the outcome of analyses of variance (ANOVAs) in SPSS 16.0 via the formula, $\rho = (MS_b - MS_w)/(MS_b + (k - 1)MS_w)$, in which ρ = ICC, MS_b = mean between-group variance, MS_w = mean within-group variance, and k = mean observations per group (Shrout & Fleiss, 1979). ICCs range from 0 to 1, with higher values indicating greater proportions of between-level variance and with an inherent greater likelihood of bias if the multilevel nature of the data is not taken into account. Multilevel modeling is warranted if ICCs are above 0.05 (Dyer et al., 2005).

Step 2: Within-group factor structure. Usually CFAs were based on the total covariance matrix (S_T). In the second step, we dissect the between-group variance from the within-group variance. The data used for analysis of the factor structure in Step 2 are in the form of the sample within-group covariance matrix, S_{PW}. The values in the S_{PW} matrix are adjusted for between-group differences by subtracting relevant group means from individual scores. If there is considerable nested variance, then the model estimated using S_{PW} may show an improved fit compared with the model estimated using S_T. The factor loadings resulting from Step 2 are usually lower than from conventional CFAs when there is substantial nested variance (Kreft & De Leeuw, 1998; Muthén, 1994), which may indicate a weaker construct validity.

Step 3: Between-group factor structure. In the third step, we investigated if the factor structure was stable across the affected and the control group (Dyer et al., 2005; Muthén, 1994). This analysis is based on the between-group covariance matrix, S_B (the covariance matrix of observed group means adjusted for the grand mean). Poor fit indices would point toward the lack of a robust factor structure across groups.

Step 4: MCFA. If the previous steps have shown that the construct validity is not stable across groups, then a MCFA is warranted to test whether the factor structure at the within-group level is robust at the between-group level (Dyer et al., 2005; Grilli & Rampichini, 2007; Muthén, 1994). Poor fit indices would indicate a significant difference between the within-group level and the between-group level. This would imply that the factor structure would not be stable at the group level and that mental health concepts are not comparable across groups. LISREL 8 was used to perform these analyses (Jöreskog & Sörbom, 1993).

Results

Analyses of demographic characteristics of the two samples revealed only differences in the distribution of the variable religion, $\chi^2(1) = 43.16$; $p < .001$. The individual responses to the items showed no substantial skewness, kurtosis, or outliers.

The affected group scored significantly higher than the control group on the items of the subscales of anxiety, $M = 2.52$, $SD = .63$; and $M = 1.92$, $SD = .67$; respectively; $t(622) = 11.43$, $p < .001$, and depression, $M = 2.48$, $SD = .40$; and $M = 1.89$, $SD = .56$; respectively; $t(622) = 13.77$, $p < .001$; see Table 1.

Construct Validity

Three defined factor models were tested using the total sample matrix (see Table 2). The orthogonal two-factor model resulted in a loss of fit compared with the one-factor model ($\Delta\chi^2 = -235.54$, $p < .001$) for the same amount of degrees of freedom. The oblique two-factor model fits the data significantly better than the one-factor model, $\Delta\chi^2(1) = 245.48$, $p < .001$. These findings indicate that the oblique two-factor structure fit the data better than the one-factor model and the two-noncorrelated-factor model.

The first column in Table 3 shows the standardized loadings of the oblique two-factor model. All of the presented estimates of the factor loadings as results of the CFA are significant, meaning that all ratios of the estimated values and their standard errors had a t value >1.96, thus $p < .05$.

Table 4 summarizes the findings of multisample CFAs to test the robustness of the factor structure across the affected and nonaffected groups. All p values of the χ^2 comparisons of the factor structure models between the affected sample and the nonaffected sample were significant at the .001 level. All hypotheses of factorial invariance were rejected. These findings indicate that the factor structure across samples is not stable, so that the constitution of the theoretical concept (latent factor) is not the same in the different samples.

Table 2. *Fit Indices of the Assessment of the HSCL-25 for Predetermined Factor Models on the Basis of Conventional CFAs*

Models	χ^2	df	CFI	RMSEA (90% Confidence Interval)
One-factor model	1,253.04*	252	0.95	0.089 (0.085–0.093)
Orthogonal two-factor model	1,488.58*	252	0.94	0.083 (0.079–0.088)
Oblique two-factor model	1,017.66*	251	0.96	0.073 (0.069–0.078)

* $p < .001$.

Table 3. *Standardized Factor Loadings From the CFA and Step 2 and 3 in the MCFA*

	Standardized Loadings					
	Total (CFA)		Within-Group Factor Structure (Step 2 in MCFA)		Between-Group Factor Structure (Step 3 in MCFA)	
Item	Factor 1	Factor 2	Factor 1	Factor 2	Factor 1	Factor 2
1	.69		.21		.26	
2	.67		.23		.28	
3	.75		.44		.29	
4	.81		.50		.23	
5	.80		.39		.29	
6	.76		.22		.34	
7	.76		.46		.18	
8	.58		.25		.19	
9	.68		.23		.14	
10	.59		.29		.13	
11		.62		.32		.34
12		.47		.13		.26
13		.56		.22		.33
15		.44		.20		.21
16		.66		.36		.35
17		.66		.44		.31
18		.67		.39		.33
19		.50		.26		.27
20		.37		.08		.29
21		.36		.01		.25
22		.68		.41		.34
23		.57		.30		.28
24		.69		.40		.34
25		.67		.32		.37

Step 1: Nested variance. The ICCs were calculated for the subscales of anxiety and depression to determine the extent of systematic group-level variance. Our data are clustered at two

Table 4. *Test of the Equality of Factor Structures of the HSCL-25 Among the Disaster-Affected and Nonaffected Sample*

Model	$\chi^2_{Control}$	$\chi^2_{Affected}$	$\Delta\chi^2$	Δdf
A	948.02	906.95	—	—
B	994.37	929.74	69.24	23*
C	1,116.50	999.55	191.94	25*
D	1,159.22	1,025.17	68.34	2*

Note. Model A = model in which the number and pattern of factors are equal across samples; Model B = Model A with the additional constraint that the factor loadings are equal across samples; Model C = Model B with the additional constraint that the error variances are equal across samples; Model D = Model C with the additional constraint that the covariance matrices of factors are equal across samples. $\chi^2_{Control,\ Affected} = \chi^2$ value of tested factor structure (model) in sample; $\Delta\chi^2 = \chi^2$ difference between two hierarchical models of invariance; Δdf = difference in degrees of freedom between two models of invariance.

* $p < .001$.

Table 5. *Model Fit of Single and Multilevel Factor Structures*

Models	χ^2	*df*	CFI	RMSEA (90% Confidence Interval)
CFA: Total	1,017.66	251	0.96	0.068 (0.064–0.073)
MCFA Step 2: Within	1,088.14	251	0.80	0.067 (0.069–0.078)
MCFA Step 3: Between	7,154.40	251	0.91	0.023 (0.022–0.023)
MCFA Step 4: Multilevel	2,114.53	527	0.51	0.060 (0.055–0.065)

Note. χ^2 = value tested factor structure in sample; $\Delta\chi^2 = \chi^2$ difference between two hierarchical models of invariance; Δdf = difference in degrees of freedom between two models of invariance.
* $p < .001$.

levels: (a) at the Gram Panchayat level because of clustered sampling, and (b) at the group level (affected vs. control group). At the Gram Panchayat level, the ICC values ranged between 0.01 and 0.04, which indicated negligible nested variance. On the group level (the affected and the control group), the ICC was .29 for the subscale of anxiety and .49 for the subscale of depression. These high ICC values indicate that data are nested on the group level; therefore, we have to specify the between-group covariance matrix (S_B) in the multilevel analyses to follow.

Step 2: Within-group factor structure. The within-group factor structure in Step 2 showed a slightly worse fit in comparison to the original CFA for the same amount of degrees of freedom ($\Delta\chi^2 = -70.48$; Step 2 RMSEA = .067). The CFI in this step indicates worse fit than for the original CFA (CFI = .080). Fan, Thompson, and Wang (1999) explain that the CFI is not effective if most of the correlations between observed variables approach 0 because there is, therefore, less covariance to explain. Such low correlations are reflected in the low factor loadings that follow. The results from Step 2 are displayed in Table 5.

The factor loadings for this within-group model in Step 2 were substantially lower than Step 1 and ranged between .21 and .50 for anxiety and between .01 and .44 for depression (see Column 2 in Table 3). Specifically, item 1 (*suddenly scared for no reason*), item 2 (*feeling fearful*), item 6 (*trembling*), item 8 (*headaches*), item 9 (*spells of terror or panic*), and item 10 (*feeling restless and can't sit still*) have very low loadings on the factor anxiety. Further, item 12 (*blaming yourself for things*), item 13 (*crying easily*), item 15 (*poor appetite*), item 19 (*feeling lonely*), item 20 (*thoughts of ending your life*), and item 21 (*feeling of being trapped or caught*) have very low loadings on the factor depression. Thus, in contrast to the results of the conventional CFA, the results of Step 2 show that the factor structure (i.e., robustness of the construct) of the subscales of depression and anxiety is poor.

Step 3: Between-group factor structure. The χ^2 in Table 5 shows that the oblique two-factor model has substantially poorer fit with the within-group covariance matrix. The χ^2 value is much larger than seen in the conventional CFA for the same amount of degrees of freedom, $\Delta\chi^2(252) = -6{,}066.26$, CFI = .91, RMSEA = .023; and Step 2, $\Delta\chi^2(251) = -4{,}968.27$. The results point toward the lack of robustness of the factor structure across the affected and control group (see column "Between-Group Factor Structure" in Table 3).

Step 4: MCFA. The results of the MCFA showed that the factor structure lost its robustness at the group level when constraining the factor loadings and the factor correlation to be invariant across the individual and group level, $\chi^2(527) = 2{,}114.53$, $p < .001$; CFI = .51; RMSEA = .060; see Table 5. Consequently, the constructs of anxiety and depression are not comparable across the affected group and the control group. The multilevel (co)variance structure path diagram of the HSCL-25 is displayed in Figure 1.

Discussion

This article aimed to illustrate methodological consequences that stem from the effect of the disaster context on screening outcomes. Similar to other disaster studies, the initial comparison of mean scores on HSCL-25 subscales for anxiety and depression showed that the disaster-affected group scored much higher on anxiety and depression than the nonaffected group. However, the difference in internal consistency of anxiety and depression in the affected group and in the nonaffected group is a first indication of a specific type of bias. We consistently demonstrated two methodological problems that hampered the comparison of mental health screening scores across these groups. The relevance of the study findings is that these problems are likely to have equally plagued other disaster studies.

First, we hypothesized that because disasters typically cause great material destruction and a loss of social capital (Kawachi & Subramanian, 2006), the disaster context evokes increased interdependence among individual mental health outcomes within an affected population (cf., Killip, Mahfoud, & Pearce, 2007). The confirmation of this hypothesis expressed itself as a nested variance problem across the group from a disaster context and the group from an unaffected environment. The problem of nested variance across the affected and nonaffected group was excessive in our dataset. This means that the scores on the HSCL-25 subscales of anxiety and depression we found in the disaster-affected group were biased because part of the covariance between observed mental health outcomes can be ascribed to the effect of the disaster context on the assessment (Kreft & De Leeuw, 1998). Not accounting for this nested variance violates the assumptions of most statistical analyses, including our initial comparison of means (Muthén, 1994).

Second, many authors have warned that the factor structure of constructs may vary across different measurement levels of data (Bliese & Hanges, 2004; Dyer et al., 2005; Härnqvist, 1978; Muthén, 1994). Indeed, the results of the multisample CFA showed that the concepts (i.e., factor structures) of anxiety and depression differed across the affected and unaffected group. The findings from the MCFA further revealed that group-level variance

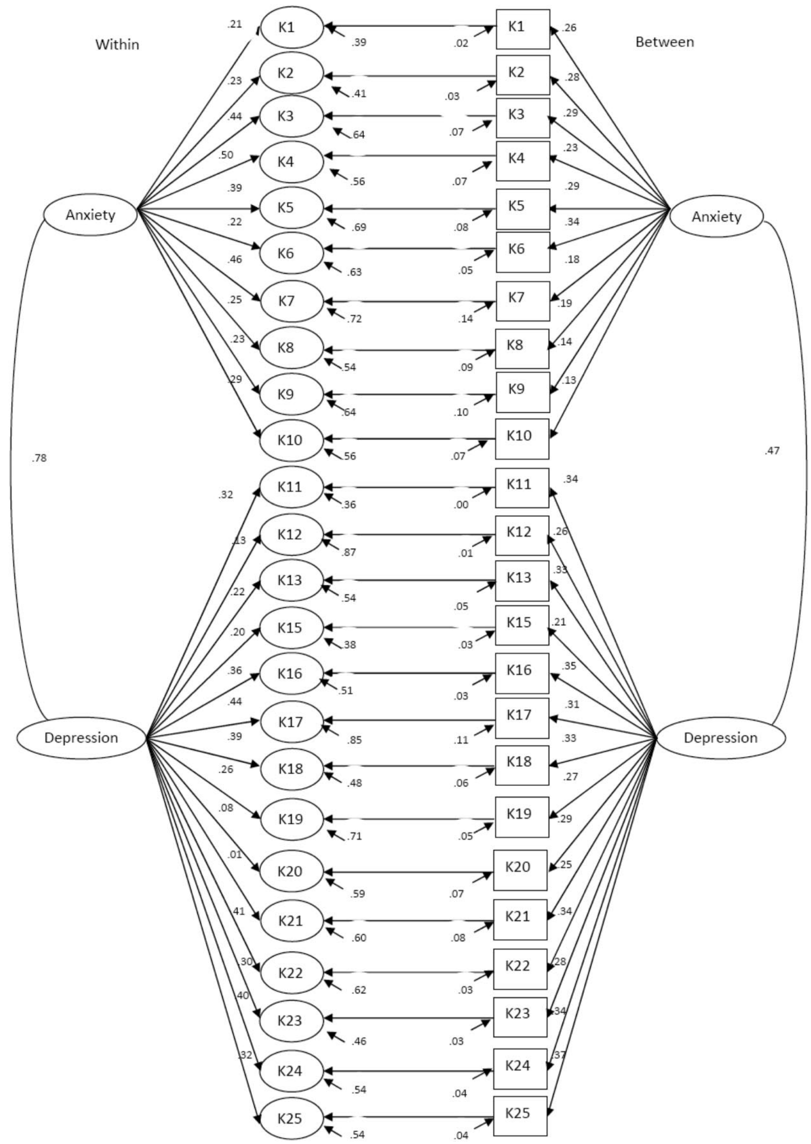

Figure 1. Multilevel (co)variance structure path diagram of the HSCL-25 divided for within-group covariance and between-group covariance (disaster affected vs. nonaffected group): latent variables, observed variables, and random error terms (i.e., unexplained variance).

(i.e., nested variance) masked the actual low and unstable construct validity of anxiety and depression. This was shown by the fit indices of the model that relied on individual within-group variance compared with the lower fit of models in which between-group variance was unaccounted for (i.e., the CFA and the MCFA). This better fitting and statistically more accurate model that is based on within-group covariance (i.e., construct-relevant covariance) showed low factor loadings and weak construct validity of the subscales of anxiety and depression in our study.

The two illustrated problems—respectively the presence of group-level variance and the difference in conceptual domains across groups (i.e., poor and unstable factor structures)—impede the comparison of both constructs across the affected and control group. In addition, both methodological problems contribute to the differences in mean scores across groups that we initially found. Namely, the difference in scores across both groups refers to differences in contexts and in concepts rather than to differences in mental health scores across groups.

These two methodological problems are not limited to our study. Most disaster mental health research relied on screening outcomes because of the practical applicability of screening instruments (Connor, Foa, & Davidson, 2006). In addition, given the fact that the destructive effect on the context is an inherent part of catastrophic events, it is likely that in other studies part of the covariance in individual mental health scores is also related to the postdisaster context rather than to the latent mental health concept per se. Or worse, disaster studies may have compared unequal constructs just as in our study. The problem of comparing unequal constructs cannot be dismissed without the required analyses (Poortinga, 1975). However, the demonstrated multilevel analyses in this article (Muthén, 1994) have not been applied in disaster mental health research; therefore, the size of the problems is hard to estimate.

The methodological problems we illustrated in this study ultimately refer to the topic that was put forth by Horwitz (2007): Whether specific mental health symptoms may constitute actual mental health problems depends on the context in which they occur. Namely, anxiety (e.g., being on high alert to danger, tension, and fear) may be an adequate reaction in an environment that has recently been hit by a flood and may be struck again. However, these same symptoms may be an inadequate reaction in an unaffected context, and in such a context, these symptoms may represent the actual "stand alone" individual mental health problems that screening instruments intend to measure. Accordingly, after we extracted the covariance related to the context, we found a factor structure with estimates that suggest that the latent constructs of the expressed mental health problems were different across both groups. In other words, expressed symptoms of mental health problems indeed depend on the context.

The study is beset by some limitations. First, both samples differ significantly on religion, because the affected group comprises fewer Muslims. Religion may have created an additional source of between-level (co)variance of the HSCL-25 items in our study. Thus, apart from the difference in context (disaster or nonaffected), religion may have partly influenced our analyses on the between-level (MCFA). Second, the HSCL-25 has not been validated in India. Nonetheless, the HSCL-25 has been previously used in the same region of Uttar Pradesh, India (Wind et al., 2013) and in the vicinity (among Tibetans in India; Crescenzi et al., 2002). Moreover, the translation procedure described in the *Method* section was thorough and accurate. Furthermore, the questionnaire was part of a larger interview that was administered by extensively trained local surveyors (students at the University of Delhi). Therefore, we assume that a possible systematic bias as a result of the lack of validation would have influenced the outcomes of the disaster group and the nonaffected group.

Despite some limitations, the results of this study contribute to the explanations for the wide range of outcomes when mental health outcomes in disaster studies are compared (Rodin & van Ommeren, 2009). The study identified, discussed, and illustrated conceptual issues and related analyses that could be highly informative to researchers working in the disaster area, especially when it concerns the estimation of the effect of the disaster context on mental health screening scores. We illustrated a procedure how to examine the construct validity of a screenings tool that is masked by nested variance as a result of the disaster (Muthén, 1994). Herewith, the study is an invitation to apply this MCFA procedure and reveal the extent of both problems in disaster mental health research. As such, scholars may determine to what extent both problems account for variation in mental health screening outcomes across disaster studies. The application of MCFA (Dyer et al., 2005; Muthén, 1994) and the identification of ecological variables that account for nested variance (Wind & Komproe, 2012) will advance our understanding of mental health in disaster contexts (Kawachi & Subramanian, 2006).

Keywords: victims of disaster; trauma; anxiety; depression; multilevel confirmatory factor analysis

References

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin, 107,* 238–246. doi:10.1037/0033-2909.107.2.238

Bliese, P. D., & Hanges, P. J. (2004). Being too liberal and too conservative: The perils of treating grouped data as though it is independent. *Organizational Research Methods, 7,* 400–417. doi:10.1177/1094428104268542

Bollen, K. A. (1991). *Structural equations with latent variables*. New York, NY: Wiley Interscience Publications.

Brown, T. (2006). *Confirmatory factor analysis for applied research*. New York, NY: The Guilford Press.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136–162). Newbury Park, CA: Sage.

Byrne, B. M., Shavelson, R. J., & Muthén, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin, 105,* 456–466. doi:10.1037/0033-2909.105.3.456

Connor, K. M., Foa, E. B., & Davidson, J. R. (2006). Practical assessment and evaluation of mental health problems following a mass disaster. *Journal of Clinical Psychiatry, 67*(Suppl. 2), 26–33.

Crescenzi, A., Ketzer, E., van Ommeren, M., Phuntsok, K., Komproe, I., & de Jong, J. T. (2002). Effect of political imprisonment and trauma history on recent Tibetan refugees in India. *Journal of Traumatic Stress, 15,* 369–375. doi:10.1023/A:1020129107279

Derogatis, L. R., Lipman, R. S., Rickels, K., Uhlenhuth, E. H., & Covi, L. (1974). The Hopkins Symptom Checklist (HSCL): A self-report symptom inventory. *Behavioral Science, 19,* 1–15. doi:10.1002/bs.3830190102

Devins, G. M., Orme, C. M., Costello, C. G., Binik, Y. M., Frizzel, B., Stam, H. J., & Pullin, W. M. (1988). Measuring depressive symptoms in illness populations: Psychometric properties of the Center for Epidemiologic Studies Depression (CES-D) scale. *Psychology & Health, 2,* 139–156. doi:10.1080/08870448808400349

Dyer, N. G., Hanges, P. J., & Hall, R. J. (2005). Applying multilevel confirmatory factor analysis techniques to the study of leadership. *The Leadership Quarterly, 16,* 149–167. doi:10.1016/j.leaqua.2004.09.009

Fan, X. B., Thompson, B., & Wang, L. (1999). Effects of sample size, estimation method, and model specification on structural equation modeling fit indexes. *Structural Equation Modeling, 6,* 56–83. doi:10.1080/10705519909540119

Galea, S., Maxwell, A. R., & Norris, F. (2008). Sampling and design challenges in studying the mental health consequences of disasters. *International Journal of Methods in Psychiatric Research, 17*(Suppl. 2), S21–S28. doi:10.1002/mpr.267

Grilli, L., & Rampichini, C. (2007). Multilevel factor models for ordinal variables. *Structural Equation Modeling, 14,* 1–25.

Härnqvist, K. (1978). Primary mental abilities of collective and individual levels. *Journal of Educational Psychology, 70,* 706–716. doi:10.1037/0022-0663.70.5.706

Horwitz, A. V. (2007). Transforming normality into pathology: The DSM and the outcomes of stressful social arrangements. *Journal of Health and Social Behavior, 48,* 211–222. doi:10.1177/002214650704800301

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. doi:10.1080/10705519909540118

Irvine, S. H., & Caroll, W. K. (1980). Testing and assessment across cultures: Issues in methodology and theory. In H. Triandis (Ed.), *The handbook of cross-cultural psychology* (pp. 127–180). Boston, MA: Allyn & Bacon.

Jöreskog, K., & Sörbom, D. (1993). *Lisrel 8: Structural equation modelling with the SIMPLIS command language*. Chicago, IL: Scientific Software International.

Kawachi, I., & Subramanian, S. V. (2006). Measuring and modeling the social and geographic context of trauma: A multilevel modeling approach. *Journal of Traumatic Stress, 19,* 195–203. doi:10.1002/jts.20108

Kessler, R. C., & Wittchen, H. U. (2008). Post-disaster mental health need assessment surveys—The challenge of improved future research. *International Journal of Methods in Psychiatric Research, 17*(Suppl. 2), S1–S5. doi:10.1002/mpr.266

Killip, S., Mahfoud, Z., & Pearce, K. (2004). What is an intracluster correlation coefficient? Crucial concepts for primary care researchers. *Annals of Family Medicine, 2,* 204–208. doi:10.1370/afm.141

Kreft, I., & De Leeuw, J. (1998). *Introducing multilevel modeling*. Thousand Oaks, CA: Sage.

Leon, G. R. (2004). Overview of the psychosocial impact of disasters. *Prehospital and Disaster Medicine, 19,* 4–9,

Lipman, R. S., Covi, L., & Shapiro, A. K. (1979). The Hopkins Symptom Checklist (HSCL)—Factors derived from the HSCL-90. *Journal of Affective Disorders, 1,* 9–24. doi:10.1016/0165-0327(79)90021-1

Muthén, B. O. (1990). *Mean and covariance structure analysis of hierarchical data*. Paper presented at the Psychometric Society, Princeton, NJ.

Muthén, B. O. (1994). Multilevel covariance structure analysis. *Sociological Methods and Research, 22,* 376–398. doi:10.1177/0049124194022003006

Poortinga, Y. H. (1975). Limitations on intercultural comparison of psychological data. *Nederlands Tijdschrift voor de Psychologie en haar Grensgebieden, 30,* 23–39.

Rodin, D., & van Ommeren, M. (2009). Commentary: Explaining enormous variations in rates of disorder in trauma-focused psychiatric epidemiology after major emergencies. *International Journal of Epidemiology, 38,* 1045–1048. doi:10.1093/ije/dyp203

Scott, R. L., Knoth, R. L., Beltran-Quiones, M., & Gomez, N. (2003). Assessment of psychological functioning in adolescent earthquake victims in Colombia: Using the MMPI-A. *Journal of Traumatic Stress, 16,* 49–57. doi:10.1023/A:1022011427985

Shrout, P. E., & Fleiss, I. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin, 86,* 420–428. doi:10.1037/0033-2909.86.2.420

van Ommeren, M., Sharma, B., Thapa, S., Makaju, R., Prasain, D., Bhattarai, R., & de Jong, J. V. T. M. (1999). Preparing instruments for transcultural research: Use of translation monitoring form with Nepali-speaking Bhutanese refugees. *Transcultural Psychiatry, 36,* 285–301. doi:10.1177/136346159903600304

Wind, T. R., Fordham, M., & Komproe, I. H. (2011). Social capital and post-disaster mental health. *Global Health Action, 4*. Retrieved May 11, 2012 from http://www.globalhealthaction.net/index.php/gha/article/view/6351.

Wind, T. R., Joshi, P. C., Kleber, R. J., & Komproe, I. H. (2013). The impact of recurrent disasters on mental health: A study on seasonal floods in northern India. *Prehospital and Disaster Medicine, 28,* 279–285. doi:10.1017/S1049023X13000290

Wind, T. R., & Komproe, I. H. (2012). The mechanisms that associate social capital with disaster mental health: A multilevel model. *Social Science & Medicine, 75,* 1715–1720. doi:10.1016/j.socscimed.2012.06.032

Winokur, A., Winokur, D. F., Rickels, K., & Cox, D. S. (1984). Symptoms of emotional distress in a family planning service: Stability over a four-week period. *The British Journal of Psychiatry, 144,* 395–399. doi:10.1192/bjp.144.4.395

World Medical Assembly Declaration of Helsinki. (1997). Recommendations guiding physicians in biomedical research involving human subjects. *Journal of the American Medical Association, 277,* 925–926. doi:10.1001/jama.1997.03540350075038

Yehuda, R., & McFarlane, A. C. (1995). Conflict between current knowledge about posttraumatic stress disorder and its original conceptual basis. *The American Journal of Psychiatry, 152,* 1705–1713.