Psychological Medicine (2013), **43**, 1611–1623. © Cambridge University Press 2013
doi:10.1017/S0033291712002772

ORIGINAL ARTICLE

Modeling and treating internalizing psychopathology in a clinical trial: a latent variable structural equation modeling approach

M. G. Kushner[1]*, R. F. Krueger[2], M. M. Wall[3], E. W. Maurer[1], J. S. Menk[4] and K. R. Menary[1]

[1] *Department of Psychiatry, University of Minnesota, Minneapolis, MN, USA*
[2] *Department of Psychology, University of Minnesota, Minneapolis, MN, USA*
[3] *Departments of Psychiatry and Biostatistics, Columbia University, New York City, NY, USA*
[4] *Biostatistical Design and Analysis Center, Clinical and Translational Science Institute, University of Minnesota, Minneapolis, MN, USA*

Background. Clinical trials are typically designed to test the effect of a specific treatment on a single diagnostic entity. However, because common internalizing disorders are highly correlated ('co-morbid'), we sought to establish a practical and parsimonious method to characterize and quantify changes in a broad spectrum of internalizing psychopathology targeted for treatment in a clinical trial contrasting two transdiagnostic psychosocial interventions.

Method. Alcohol dependence treatment patients who had any of several common internalizing disorders were randomized to a six-session cognitive-behavioral therapy (CBT) experimental treatment condition or a progressive muscle relaxation training (PMRT) comparison treatment condition. Internalizing psychopathology was characterized at baseline and 4 months following treatment in terms of the latent structure of six distinct internalizing symptom domain surveys.

Results. Exploratory structural equation modeling (ESEM) identified a two-factor solution at both baseline and the 4-month follow-up: Distress (measures of depression, trait anxiety and worry) and Fear (measures of panic anxiety, social anxiety and agoraphobia). Although confirmatory factor analysis (CFA) demonstrated measurement invariance between the time-points, structural models showed that the latent means of Fear and Distress decreased substantially from baseline to follow-up for both groups, with a small but statistically significant advantage for the CBT group in terms of Distress (but not Fear) reduction.

Conclusions. The approach demonstrated in this study provides a practical solution to modeling co-morbidity in a clinical trial and is consistent with converging evidence pointing to the dimensional structure of internalizing psychopathology.

Received 21 December 2011; Revised 1 November 2012; Accepted 4 November 2012; First published online 9 January 2013

Key words: Anxiety disorder, clinical trial, internalizing, latent variable structural equation modeling, psychopathology.

Introduction

Since DSM-III was introduced (APA, 1980), internalizing psychopathology has been parceled into multiple disorder subtypes that are presumed to be etiologically distinct and otherwise independent [e.g. generalized anxiety disorder (GAD), social anxiety disorder (SAD), major depressive disorder (MDD) and panic disorder (PD)]. However, this model has difficulty accommodating the high inter-correlation among various anxiety and depressive disorders (e.g. Beekman *et al*. 2000; Brown *et al*. 2001; Andrews *et al*. 2002; Kessler *et al*. 2005; Kushner *et al*. 2005; Moffitt *et al*. 2007) and does not correspond to the substantial and ever-growing corpus of published scientific work pointing to the dimensional structure of internalizing psychopathology (e.g. Krueger, 1999; Kendler *et al*. 2003). For example, studies in epidemiological psychiatry have demonstrated that the empirical structure (i.e. the patterns of covariation) of up to seven common internalizing disorder diagnoses can be modeled by a single higher-order latent variable ('trait') commonly labeled as 'Internalizing' (e.g. Krueger, 1999). This higher-order factor has also been found to be composed of two lower-order factors commonly labeled as 'Fear' (SAD, agoraphobia, PD and simple phobia) and 'Distress' (GAD, MDD and dysthymia) (e.g. Krueger, 1999; Andrews *et al*. 2009). This factor

* Address for correspondence: M. G. Kushner, Ph.D., 282-2A West, 2450 Riverside Avenue, Minneapolis, MN 55454, USA.
(Email: kushn001@umn.edu)

structure for internalizing disorders has proven highly replicable across multiple epidemiological studies using community resident respondents (cf. Krueger & Markon, 2006) and has also been found in psychiatric out-patients (McGlinchey & Zimmerman, 2007). Based on this, we sought to establish an analytic strategy that would accommodate the dimensional structure of internalizing psychopathology in a clinical trial designed to contrast the treatment effects of two transdiagnostic psychosocial interventions.

Consistent with the neo-Kraepelinian framework upon which the modern psychiatric diagnostic nomenclature is built, cognitive-behavioral therapy (CBT) protocols (along with many other types of therapy) have been designed and promulgated primarily for the treatment of a specific internalizing disorder subtype such as MDD (e.g. Beck *et al*. 1979), SAD (Heimberg *et al*. 1990) or PD (Barlow *et al*. 1989). However, pervasive correlated disorders undermine the efficiency, if not the logic, of clinical interventions narrowly targeted to a single internalizing disorder. Therefore, we adapted and integrated a variety of standard CBT techniques to target the more general internalizing psychopathology domains of Fear (e.g. using exposure therapy; Foa & Kozak, 1986) and Distress (e.g. using cognitive restructuring; Ellis & Harper, 1975; Beck, 1976). In this study, we contrasted the resulting CBT program with progressive muscle relaxation training (PMRT; Jacobson, 1938; Bernstein & Borkovec, 1973). PMRT was chosen as the contrast condition because: (1) PMRT provides a real-world 'effectiveness' test of CBT compared to standard treatment for internalizing problems (i.e. relaxation training); (2) PMRT controls numerous threats to the study's validity because it is typically viewed as a compelling and reasonable treatment for internalizing problems; (3) PMRT is highly structured and manualized so that sessions are reliable from subject to subject; and (4) PMRT, although narrowly targeted to teaching patients how to relax their muscles, is not diagnostically specific; that is, PMRT, like CBT, is a transdiagnostic intervention.

Because the latent modeling of internalizing psychopathology identified in previous work was largely based on lifetime psychiatric diagnoses as manifest indicators (e.g. Krueger, 1999), that approach would not be sensitive to treatment effects; that is, symptom reduction due to treatment would not be reflected in a change in a person's lifetime diagnostic status. However, if the established covariation pattern of common depression and anxiety diagnoses results from the influence of a latent internalizing psychopathology trait (as we believe it does), then the empirical structure of internalizing symptoms should correspond to that of the internalizing diagnoses. In fact, Markon (2010) and Simms *et al*. (2012) demonstrated that the actual symptoms making up the various anxiety and depressive diagnoses manifest roughly the same latent empirical structure as do the diagnoses themselves. Based on these considerations and past findings, we assessed manifest internalizing psychopathology before and after treatment using validated dimensional measures targeted to specific anxiety and depression domains (i.e. social anxiety, depression, generalized anxiety/worry, agoraphobia, panic anxiety and trait anxiety). As noted, we expected to find the same basic Fear and Distress latent subdimensions of the broader internalizing spectrum that were often, but not universally (see, for example, Kessler *et al*. 2011), found in earlier work (Krueger & Markon, 2006; Watson, 2009; Eaton *et al*. 2011; Kushner *et al*. 2012; Simms *et al*. 2012).

The first step in the analytic strategy entailed exploratory structural equation modeling [ESEM, a modern improvement over exploratory factor analysis (EFA)] to ascertain the empirical (covariation) structure of the manifest internalizing psychopathology symptom domain measures at baseline and at follow-up and then evaluate whether these structures were similar at both time points. This step was meant to confirm that the relationship among the measures (as opposed to the level/intensity of symptoms) remained constant before and after treatment. Next we examined the difference in the change in the means (intercepts) of the latent factor(s) from before to after treatment between the two groups. This step is meant to characterize the overall reduction in internalizing psychopathology in the sample from before to after treatment and between the two study groups on the level of the latent internalizing psychopathology variable(s). Finally, we examined whether the means (intercepts) of the six manifest indicators changed differentially from pre- to post-treatment relative to the latent variable(s). This final step allows us to determine the extent to which the rate of change seen within and between the groups is captured completely by the latent factor(s), or whether there is additional information about treatment response in components of the manifest symptom measures that is not captured by the latent factor(s). We are unaware of any published literature using these methods or pursuing this research agenda in a clinical trial.

Method

The data used in this study were collected in a multi-year project investigating interventions aimed at improving alcohol dependence treatment outcomes in patients with co-occurring anxiety disorders. Descriptions of recruitment, retention and the treatments

in this report are structured to its limited scope and aims relative to those of the primary project in which these data were collected (cf. Kushner *et al.* in press). The interested reader can find a broader description of the study, along with results related to alcohol outcomes (*versus* the internalizing psychopathology outcomes that are the focus of this report), upon request to the corresponding author.

Participants

Inclusion/exclusion criteria

All participants were drawn from an adult (age >18 years) community-based residential alcohol dependence treatment program located in a major metropolitan area in the Upper Midwest region of the USA. Inclusion required meeting diagnostic criteria for current (past 30 days) alcohol dependence and at least one of the following qualifying anxiety disorders: PD, SAD or GAD. Individuals were excluded from participation if they had a history of bipolar disorder or schizophrenia or if they had conditions deemed likely to interfere with their capacity to fully participate in the study, e.g. cognitive impairment, serious ongoing suicidality or inability to read and understand English. Patients were not excluded if they had MDD, posttraumatic stress disorder (PTSD) or drug (other than alcohol) dependence. However, we only included those for whom alcohol dependence (rather than drug dependence) was their primary reason for seeking treatment. This study was approved by the Human Subjects Committee of the University of Minnesota's Institutional Review Board (IRB). Each participant provided informed consent to participate in the study.

Recruitment

Screening for qualified participants involved three levels of assessment that took place within a patient's first week of treatment in the 21-day alcohol use disorder (AUD) treatment program. First, we offered a brief self-report screening questionnaire to all of the patients entering the AUD treatment program, asking patients to indicate their primary substance of abuse and whether they had experienced any disturbing 'anxiety attacks', excessive worry or anxiety/discomfort in social situations in the past 30 days. Those who seemed to be qualified were then asked by a trained research assistant to elaborate on the endorsements they made in response to the earlier screen questions. Status on the exclusion criteria were also assessed at this time. Based on the interviewer's notes, the clinical team (including at least one staff Ph.D.-level psychologist) decided whether individuals should be invited for the third and final screening step, which included the Structured Clinical Interview for DSM-IV (SCID; First *et al*. 1989), to confirm the presence of inclusionary diagnoses.

Of the 344 individuals who were randomized in the study ($n=171$ in CBT and $n=173$ in PMRT), SEM analyses were conducted using the 322 cases that had completed at lease one of the baseline internalizing assessments ($n=159$ in CBT and $n=163$ in PMRT). (See assumptions and analytic steps related to cases with missing data at the follow-up for SEM analyses described below.) However, only cases that had completed both baseline and follow-up internalizing assessments were included in the ancillary ANCOVAs described below ($n=239$–241 for these analyses).

Interventions

CBT: experimental treatment

The CBT program splits its six 1-h sessions into three primary content domains (psycho-education, cognitive restructuring and exposure/habituation), each of which are the focus of two yoked sessions: one focused on anxiety symptoms exclusively and one focused on the interaction between anxiety symptoms and alcohol use. The content for the six sessions was synthesized by M.G.K. from a survey of published literature related to CBT for anxiety disorders, alcohol disorders and the association between alcohol use and anxiety disorders (cf. Kushner *et al.* in press). Each of the six sessions were delivered at the conclusion of the AUD treatment day (at ~15:30 hours) on six consecutive business days.

PMRT: comparison treatment

The PMRT comparison treatment was based on the treatment described in *Progressive Relaxation Training: A Manual for the Helping Professions* (Bernstein & Borkovec, 1973). The program was adapted slightly to match the CBT program in terms of session length (1 h) and session number (six). As with the CBT, PMRT sessions were delivered at the end of the AUD treatment day on six consecutive business days.

Internalizing symptom measures

Specific domains of internalizing psychopathology symptoms were assessed at baseline and at the 4-month follow-up using validated self-report symptom scales as follows: for the generalized anxiety/worry symptom domain, we used the Penn State Worry Questionnaire (PSWQ; Meyer *et al*. 1990); for the depression symptom domain, we used the Beck Depression Inventory (BDI; Beck *et al*. 1961); for the panic anxiety symptom domain, we used the Panic

https://doi.org/10.1017/S0033291712002772 Published online by Cambridge University Press

Disorder Severity Scale (PDSS; Houck *et al*. 2002); for the social anxiety symptom domain, we used the Social Phobia Scale (SPS; Brown *et al*. 1997); for the agoraphobia symptom domain, we used the Mobility Inventory for Agoraphobia (MIA; Chambless *et al*. 1985); and for the trait anxiety symptom domain, we used the Trait Anxiety version of the State–Trait Anxiety Inventory (STAI; Spielberger *et al*. 1999).

Data analytic approach

Although the covariance structure of internalizing binary diagnoses is reasonably well known (e.g. Krueger, 1999), we took an EFA approach to characterize the structure of internalizing problems in this study because our dataset was distinct from those used in the earlier studies in several potentially important ways, including: (1) we analyzed symptom surveys relevant to internalizing symptom domains rather than diagnoses themselves; (2) we assessed the structure of the internalizing measures at two time-points (before and after treatment), something that has rarely been examined; and (3) we examined the structure of the internalizing measures in a clinical (*versus* community-based) sample, something that has also rarely been examined.

Specifically, we used ESEM (Marsh *et al*. 2009), newly implemented in Mplus 6 (Muthén & Muthén, 2011), to test the factor structure (i.e. the number of factors and configuration of observed variables loading on those factors) underlying the internalizing psychopathology measures at both baseline and follow-up. ESEM integrates confirmatory factor analysis (CFA) and EFA by allowing tests of loading invariance across time, as in CFA, while permitting unrestricted EFA structure (i.e. no *a priori* fixing of cross-loadings to zero). An advantage of the integrated approach of ESEM is that it allows for assessment of configural invariance over time (i.e. invariance of which items load on which factors over time) without needing to impose a simple factor structure (i.e. only one item loading on each factor). One- and two-factor models at both time-points with and without constraining the loadings to be invariant across time were fit with ESEM using maximum likelihood and compared using the Bayesian Information Criterion (BIC). [We chose the BIC as a parameter-sensitive index over the Akaike's Information Criterion (AIC) because the former favors more parsimonious models; e.g. Vrieze, 2012.] In the model with constrained loadings across time, the scale of the latent variables was fixed to have variance equal to one at time 1 but was allowed to be freely estimated at time 2 to allow for the possibility of changing variability. An oblique geomin rotation was used because past work suggested that the factors would be correlated. In addition, models that assumed one factor at baseline and two factors at follow-up and *vice versa* were fit and compared. With only six observed variables, no more than two factors can be identified without imposing *a priori* constraints. In all models, the observed variable residual correlations across time (i.e. autocorrelations) were freely estimated. The ESEM with the smallest BIC that also satisfied typical rules of thumb governing 'good fit' [i.e. Comparative Fit Index (CFI) >0.90 and root mean squared error of approximation (RMSEA) <0.08] was taken as the candidate for further confirmatory modeling (Hu & Bentler, 1999). Using the typical rule of thumb for EFA, factor loadings found to be less than 0.3 in the ESEM were fixed to zero in subsequent CFA modeling. Finally, the ESEM was carried out using the entire sample at both time-points. This approach was taken because the randomization presumably equates the samples at baseline prior to the introduction of the interventions and also because splitting the sample in two would decrease the stability of the ESEM results.

Given our aim of testing change in the level of latent internalizing psychopathology over time and possibly differential changes over time by treatment group, it was important to verify that the latent structure was reasonably stable across time and treatment groups. We tested for strong invariance (cf. Meredith, 1993) of the measurement model by comparing a series of models constraining factor loadings and observed variable intercepts to be equivalent across time and/or treatment groups. Specifically, four models were compared: (1) 'fully constrained': factor loadings and intercepts constrained across time and group, (2) 'time constrained': free factor loadings and intercepts across groups but the same across time within group, (3) 'group constrained': free factor loadings and intercepts across time but the same across groups, and (4) 'fully unconstrained': free factor loadings and intercepts across time and groups.

Evidence supporting strong invariance comes from the fully constrained model having the best (smallest) BIC of the four models compared. In all models, the observed variable residual correlations across time (i.e. autocorrelations) were freely estimated and unconstrained across groups, as were the residual variances. Although not required for strong invariance, the factor variances were constrained to be the same at baseline and follow-up and across groups. This additional constraint facilitates a clear definition of effect sizes (i.e. change in latent factor means over time and differences in them between groups) that require a scale (variance) estimate as a denominator for calculation. Because the target variables are latent and, as such, do not intrinsically have a scale, fixing their

respective variances across time and group provides a common scaling so that effect size changes over time and between treatment groups are directly comparable and do not depend on possibly different standardizations that would arise if the variances were allowed to differ. To ensure that this additional constraint does not lead to a worse-fitting model, we also compared the BIC with a model that allows the variances to all be different.

A multiple-group (CBT *v.* PMRT treatment), two-time-point (baseline to follow-up), two-latent-factor (Distress and Fear) structural equation model was fit to the full sample to estimate the change in latent factor means over time and to estimate the potentially differential change in those means by treatment group. We fit the 'fully constrained' Distress and Fear factor model with observed variable intercepts and loadings constrained to be the same at baseline and follow-up and across the two treatment groups. The baseline factor means were fixed to zero and freely estimated at follow-up in both treatment groups. Estimates of factor mean change over time within and between treatment groups were standardized by the variance of the respective latent factors and thus represent effect sizes (across time and across treatment).

We conducted a single overall omnibus test for strong invariance; that is, whether the best-fitting model overall holds loadings and intercepts equal across time and groups. However, because of our programmatic interest in differential changes in specific observed (manifest) *versus* latent variables over time, we also conducted a series of ancillary examinations in which we fit six additional models where one of the six observed variable intercepts is allowed to be freely estimated across time. The difference in the estimated intercept across time indicates how the manifest variable changes over time compared to the latent variable it is measuring. We present data for any of the six models with better BIC than the fully constrained model.

Maximum likelihood estimation was used for all latent variable models under the assumption that all symptom measures at baseline and follow-up were normally distributed. The largest skew and kurtosis for all six symptom variables was below 3 for skew and below 10 for kurtosis, satisfying typical rules of thumb for normality assumptions used for continuous latent variable modeling (Kline, 2010). Specifically, across all six symptom variables at baseline, the range was −0.8 to 0.4 for skew and 2.4–3.2 for kurtosis, and at follow-up, the range was 0.02–1.8 for skew and 2.2–6.2 for kurtosis. The 81 cases lost to follow-up did not complete any of the six internalizing symptom measures at follow-up. Additionally, there were 16 instances of inadvertent 'missingness' of individual symptom measures (four at baseline and 12 at follow-up). Nonetheless, these cases were included using analyses computed under a missing at random (MAR) assumption using all available data in Mplus version 6.1 (Muthén & Muthén, 2011). MAR is less restrictive than the missing completely at random (MCAR) assumption used with listwise deletion and assumes missingness is not dependent on the actual value that is missing but that the value that is missing is conditional on variables we can observe about the individual (e.g. the baseline score). MAR is a common assumption made for missing data in pre–post studies where baseline variables are often predictive of missing status and can validly be modeled with latent variables (cf. Little & Rubin, 2002).

Finally, for comparison with our latent variable modeling approach, we also performed ancillary ANCOVAs to test for treatment effects for each of the six manifest internalizing symptom measures separately. Six separate models were fit (one for each internalizing measure), using those cases that completed the measure at both the baseline and the 4-month follow-up (n values for these analyses ranged between 238 and 241). Each ANCOVA included a dichotomous indicator of treatment and the respective baseline symptom measure. Effect sizes for treatment effect were taken to be the ANCOVA coefficient for treatment scaled by the standard deviation of the symptom measure at baseline.

Results

Demographic and clinical characteristics at baseline

As shown in Table 1, the groups did not differ on age, gender or qualifying anxiety disorder. Table 1 also shows that the CBT group was somewhat more symptomatic than the PMRT group on most of the internalizing anxiety and depressions measures. Although some of these effects approached statistical significance, the groups were significantly different on the BDI only ($t = 1.988$, df = 320, $p < 0.05$); however, even this effect was small in size at about 0.2 S.D. units difference between the group means. Table 2 shows the correlation matrices for the six internalizing indicators at baseline and follow-up. Not surprisingly, all correlations are positive and significant ($p < 0.001$), ranging in magnitude from 0.19 to 0.77. (Additional information related to the raw data can be obtained from the authors.)

Two-time-point ESEM

Table 3 shows the various constrained and unconstrained models generated from the two-time-point

https://doi.org/10.1017/S0033291712002772 Published online by Cambridge University Press

Table 1. *Baseline characteristics by group in the sample used for structural equation modeling (SEM) analyses*

	CBT ($n=159$)			PMRT ($n=163$)			
Variable	Mean	S.D.	%	Mean	S.D.	%	p
Age (years)	39.1	9.7		39.5	10.6		0.7145
Gender (% female)			38.4			42.3	0.4683
Major depression diagnosis			45.3			42.3	0.5935
Qualifying anxiety diagnosis[a]							0.9378
GAD			39.6			38.0	
PD			15.7			15.3	
SAD			44.7			46.6	
Symptom measures							
BDI	21.8	9.5		19.8	8.3		0.0479
STAI	58.0	10.7		56.0	10.2		0.0856
PSWQ	64.9	10.8		64.1	11.0		0.4908
MIA	1.3	0.8		1.2	0.8		0.1598
SPS	17.1	8.8		15.6	8.1		0.1330
PDSS	11.6	6.4		10.3	6.0		0.0654

CBT, Cognitive behavioral therapy; PMRT, progressive muscle relaxation therapy; GAD, generalized anxiety disorder; PD, panic disorder; SAD, social anxiety disorder; BDI, Beck Depression Inventory; STAI, State–Trait Anxiety Inventory (Trait Version); PSWQ, Penn State Worry Questionnaire; MIA, Mobility Inventory for Agoraphobia; SPS, Social Phobia Scale; PDSS, Panic Disorder Severity Scale; S.D., standard deviation.

[a] Participants meeting criteria for more than one qualifying anxiety diagnosis were asked to elect a 'principal' anxiety disorder based on the grouping of symptoms that 'troubles you the most' (Andrews *et al*. 2002). In these cases, the principal anxiety disorder was coded as the qualifying anxiety diagnosis.

Table 2. *Correlation matrix of internalizing symptom measures at baseline and 4-month follow-up*

	Baseline					
Follow-up	BDI	STAI	PSWQ	MIA	SPS	PDSS
BDI		0.621	0.308	0.322	0.354	0.358
STAI	0.765		0.430	0.364	0.428	0.362
PSWQ	0.590	0.668		0.224	0.289	0.186
MIA	0.408	0.435	0.364		0.627	0.560
SPS	0.503	0.606	0.556	0.664		0.542
PDSS	0.617	0.551	0.491	0.603	0.649	

BDI, Beck Depression Inventory; STAI, State–Trait Anxiety Inventory (Trait Version); PSWQ, Penn State Worry Questionnaire; MIA, Mobility Inventory for Agoraphobia; SPS, Social Phobia Scale; PDSS, Panic Disorder Severity Scale.

Baseline correlations between manifest indicators above the diagonal; follow-up correlations between manifest indicators below the diagonal.

All values $p<0.001$.

ESEM. The BIC fit statistic was lowest for the model with two factors at baseline and two factors at follow-up with the same configural structure at both time-points (i.e. constrained loadings). Furthermore, this two-factor model at both time-points showed good fit with the CFI and RMSEA indices (see Table 3). Table 4 shows the estimated standardized factor loadings, standard errors, and their associated p values for the two-factor solution. The same symptom measures significantly load on the same factors (Distress and Fear) when using both the baseline and follow-up measures. Note also that the correlation between the rotated Fear and Distress factors at both times 1 and 2 were moderately high, at 0.63 and 0.77 respectively. Correlation of the Fear factor over time was 0.52 whereas that for the Distress factor was smaller at 0.30. The cross-factor correlations across time were small to moderate: Fear at time 1 with Distress at time 2 = 0.20 and Distress at time 1 with Fear at time 2 = 0.25.

Measurement invariance: CFA

The BIC fit statistic was lowest (i.e. best) for the fully constrained model (BIC = 20 142) compared to the time-constrained model (BIC = 20 191), the group-constrained model (BIC = 20 148) and the unconstrained model (BIC = 20 202), indicating 'strong' invariance (i.e. invariance of factor loadings and intercepts) across time and treatment groups. Furthermore, weakening the constraint on the factor variances of the fully constrained model did not improve the BIC (= 20 160). Finally, the χ^2 test of model fit for

https://doi.org/10.1017/S0033291712002772 Published online by Cambridge University Press

Table 3. *Two-time-point exploratory structural equation model (ESEM) fit*

Model	No. of parameters	BIC	CFI	RMSEA
Unconstrained loadings across time				
One factor at BL; one factor at FU	43	20179.9	0.872	0.12
One factor at BL; two factors at FU	49	20132.3	0.917	0.103
Two factors at BL; one factor at FU	49	20133.7	0.916	0.104
Two factors at BL; two factors at FU	56	20049.9	0.985	0.048
Constrained loadings across time				
One factor at BL; one factor at FU	38	20187.5	0.853	0.122
Two factors at BL; two factors at FU	48	**20023.5**	**0.978**	**0.052**

BL, Baseline; FU, follow-up; BIC, Bayesian Information Criterion; CFI, Confirmatory Fit Index; RMSEA, root mean squared error of approximation.

Bold values indicate best-fitting model.

Table 4. *Two-time-point exploratory structural equation model (ESEM): model results for two factors with constrained loadings across time*

	Fear				Distress			
	Estimated loading	S.E.	Estimate/S.E.	*p*	Estimated loading	S.E.	Estimate/S.E.	*p*
Baseline								
BDI	0.075	0.071	1.055	0.291	**0.624**	0.061	10.175	<0.001
STAI	−0.001	0.024	−0.046	0.963	**0.869**	0.036	23.862	<0.001
PSWQ	0.060	0.076	0.787	0.431	**0.527**	0.062	8.489	<0.001
MIA	**0.947**	0.071	13.310	<0.001	−0.199	0.071	−2.784	0.005
SPS	**0.765**	0.032	23.762	<0.001	0.000	0.003	−0.121	0.904
PDSS	**0.654**	0.055	11.895	<0.001	0.039	0.052	0.748	0.455
Four-month follow-up								
BDI	0.070	0.067	1.046	0.296	**0.790**	0.065	12.176	<0.001
STAI	−0.001	0.018	−0.046	0.963	**0.919**	0.026	34.712	<0.001
PSWQ	0.053	0.068	0.778	0.437	**0.636**	0.070	9.119	<0.001
MIA	**0.992**	0.096	10.377	<0.001	−0.285	0.104	−2.748	0.006
SPS	**0.838**	0.026	32.609	<0.001	0.000	0.004	−0.121	0.904
PDSS	**0.764**	0.072	10.596	<0.001	0.062	0.082	0.752	0.452

BDI, Beck Depression Inventory; STAI, State–Trait Anxiety Inventory (Trait Version); PSWQ, Penn State Worry Questionnaire; MIA, Mobility Inventory for Agoraphobia; SPS, Social Phobia Scale; PDSS, Panic Disorder Severity Scale; S.E., standard error.

Bolded estimated factor loadings signify $p < 0.001$.

the fully constrained model was 251 on 122 degrees of freedom and it demonstrated adequate fit to the data with RMSEA = 0.081 and CFI = 0.92.

Latent variable SEM: tests of change over time and across treatment groups

The fully constrained two-group structural equation model with two factors at baseline and follow-up was used to examine the effects of time and treatment as shown in Fig. 1. The tests for the change over time within and between treatment groups for each latent factor are presented in Table 5. A significant time effect was observed in both treatment groups for both latent factors. Large improvements in Fear and Distress were observed from baseline to follow-up. The CBT group demonstrated significantly greater reduction on the Distress latent factor compared to that in the PMRT group ($p = 0.024$); however, the size of this effect was small. No statistical significant difference was found between the two groups when comparing the Fear latent factor at follow-up ($p = 0.298$).

https://doi.org/10.1017/S0033291712002772 Published online by Cambridge University Press

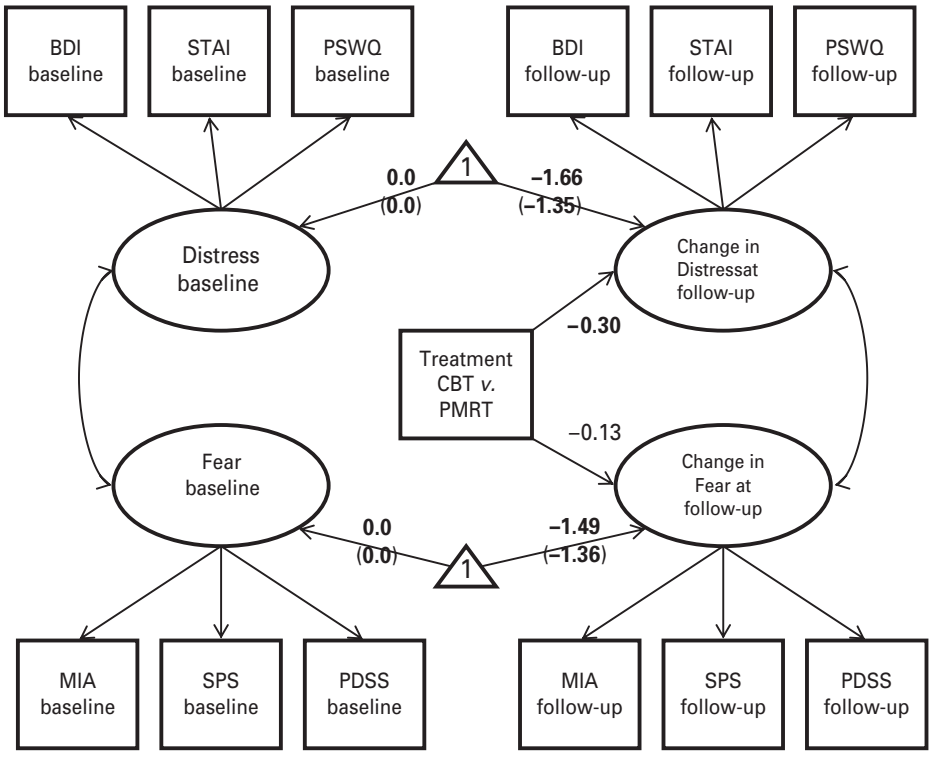

Fig. 1. Graphic structural equation model of change in latent Distress and Fear from baseline to follow-up between treatment groups. Following conventional graphical notation for latent means, estimates on arrows from triangles containing a '1' represent estimated mean values for latent factors at baseline (fixed at zero) and follow-up (i.e. effect size change from baseline); estimates above arrows are for cognitive-behavioral therapy (CBT) and below arrows are progressive muscle relaxation training (PMRT) in parentheses. Estimates leading from treatment outward represent treatment effect size; that is treatment difference in effect size change [e.g. $-0.28 = -1.98 - (-1.70)$ is the effect size of CBT *v.* PMRT]. Bolded values indicate that $p < 0.05$. Residuals for all observed symptom variables were allowed to autocorrelate across time (arrows not shown). BDI, Beck Depression Inventory; STAI, State–Trait Anxiety Inventory (Trait Version); PSWQ, Penn State Worry Questionnaire; MIA, Mobility Inventory for Agoraphobia; SPS, Social Phobia Scale; PDSS, Panic Disorder Severity Scale.

Table 5. *Treatment effects using the fully constrained two-group model*

Standardized time effects on factor means

		Change in factor mean	S.E.	Estimate/S.E.	*p* value
Fear	PMRT	−1.362	0.105	−12.916	<0.001
	CBT	−1.487	0.111	−13.421	<0.001
Distress	PMRT	−1.354	0.116	−11.674	<0.001
	CBT	−1.655	0.123	−13.408	<0.001

Standardized treatment group effects on factor means

	Group difference in change in factor mean	S.E.	Estimate/S.E.	*p* value
Fear (CBT *v.* PMRT)	−0.125	0.120	1.041	0.298
Distress (CBT *v.* PMRT)	−0.301	0.134	2.251	0.024

PMRT, Progressive muscle relaxation therapy; CBT, cognitive-behavioral therapy; S.E., standard error.

https://doi.org/10.1017/S0033291712002772 Published online by Cambridge University Press

Ancillary analyses

Change in observed variables relative to latent variables over time and treatment

Differential changes between the observed variables and the latent factors across time and group were explored using six separate models, each with one of the symptom measure intercepts unconstrained. Only one of these models had a better BIC than the fully constrained model and that was when the intercept for the PSWQ was allowed to be freely estimated at follow-up. The PSWQ symptom measure intercepts had statistically significant changes over time within each treatment. The intercept of the PSWQ at follow-up was 6.68 (S.E. = 1.04, $p < 0.0001$) units higher in the PMRT group compared to the intercept at baseline and 6.22 (S.E. = 1.16, $p < 0.0001$) units higher in the CBT group. Given the pooled standard deviation of the PSWQ at baseline of 10.9, these intercept differences correspond to effect size changes 0.61 and 0.57 higher than expected. If the PSWQ decreased over time in step with the latent distress factor, its intercept should be the same at baseline and follow-up. This change in intercept over time suggests that the PSWQ is not decreasing at the same rate as the distress factor. When comparing between groups, however, the PSWQ demonstrated no differential change over time. In other words, there was no difference between the group intercept changes on the PSWQ over time.

Change in manifest internalizing measures over time and treatment by group

Table 6 presents the results from fitting separate ANCOVA models to each of the six manifest internalizing measures; that is examining the group differences at the 4-month assessment while controlling for baseline values. All treatment effects were negative, indicating that CBT was associated with a greater decrease in symptoms than PMRT. However, this group effect was statistically significant for the STAI Trait Anxiety measure only.

Discussion

Although the high degree of inter-correlation among common internalizing disorders is well documented by epidemiological research (e.g. Krueger, 1999; Beekman *et al*. 2000; Brown *et al*. 2001; Andrews *et al*. 2002; Kessler *et al*. 2005), the implications of these and related findings for clinical research have been slow to penetrate the field. Clinical trials are typically designed to test the effect of a specific treatment on a single diagnostic entity, an approach that is in keeping with the neo-Kraepelinian assumption that each diagnostic entity is clinically independent. However, this approach does not accommodate converging evidence that the various internalizing diagnoses are observed indicators of a shared underlying internalizing trait (e.g. Mineka *et al*. 1998; Krueger, 1999; Kendler *et al*. 2003) and that the level of this trait correlates strongly with socially and clinically important outcomes (e.g. Krueger & Finger, 2001; McGlinchey & Zimmerman, 2007; Kushner *et al*. 2012). The primary aim of the present work was to establish a data analytic approach that explicitly addresses correlated internalizing disorders and their likelihood of representing an underlying internalizing psychopathology trait in a clinical trial. The cornerstone of this effort was a latent variable structural equation approach to modeling a broad range of internalizing symptom domains in a way that: (*a*) corresponds to the oft-replicated latent structure for internalizing diagnoses reported in the psychiatric epidemiology literature; (*b*) maintains this same essential latent structure over the course of time and treatment; and (*c*) produces interpretable latent variable scores sensitive to clinical change due to treatment response and time.

The latent variable SEM approach not only is more efficient from an analytic perspective (i.e. sharing information across measures to increase power) but also explicitly allows elevations on all of the covarying measures to yield empirically meaningful aggregate dimensions, creating a perspective on the data not available when looking at the manifest measures individually. In this regard, the CBT treatment was significantly more effective than the PMRT treatment in reducing the level of the latent Distress variable but not that of the latent Fear variable. Moreover, the opportunity to compare and contrast treatment effects on both the latent and manifest variables provides a unique perspective on the data not available in either approach alone. For example, ancillary analyses showed that the PSWQ decreased significantly less over time than did the latent Distress variable on which it loaded. This finding could indicate that the CBT treatment should expand its scope to better represent (i.e. treat) the domain of distress specifically related to worry as measured by the PSWQ. Alternatively, this finding might suggest that distress symptoms related to worry change at a different (i.e. slower) rate in response to treatment than do other distress symptom domains (i.e. those related to trait anxiety and depression). Ancillary analyses also showed that the STAI Trait Anxiety measure was the only one of the three manifest indicators contributing information to the latent Distress variable that was, like the Distress variable itself, significantly different between the groups. This finding suggests that Trait Anxiety was the most sensitive manifest indicator of

https://doi.org/10.1017/S0033291712002772 Published online by Cambridge University Press

Table 6. *Symptom measure means, standard deviations and effect size changes from baseline to 4-month follow-up by treatment group*

	CBT						PMRT						Comparison of treatments[c]		
		Baseline		4-month FU				Baseline		4-month FU					
	n[a]	Mean	S.D.	Mean	S.D.	Change ES[b]	*n*[a]	Mean	S.D.	Mean	S.D.	Change ES[b]	*β*	tx ES	*p* value
Distress															
BDI	111	20.72	9.06	11.20	8.20	−1.06	127	19.41	8.20	12.40	8.89	−0.78	−1.55	−0.17	0.152
STAI	112	57.93	10.35	41.49	12.33	−1.57	129	55.35	10.48	43.99	12.34	−1.09	−3.36	−0.32	0.031
PSWQ	110	64.68	10.35	47.21	13.27	−1.60	128	64.38	10.62	49.03	12.52	−1.41	−1.97	−0.18	0.201
Fear															
MIA	112	1.36	0.78	0.61	0.73	−0.93	128	1.18	0.82	0.53	0.73	−0.79	−0.02	−0.02	0.847
SPS	112	16.58	8.54	7.04	6.46	−1.12	129	15.32	8.33	7.25	7.73	−0.95	−0.77	−0.09	0.329
PDSS	111	11.43	6.42	4.24	4.60	−1.16	129	10.21	6.0	4.73	4.93	−0.88	−0.81	−0.13	0.166

CBT, Cognitive-behavioral therapy; PMRT, progressive muscle relaxation therapy; BDI, Beck Depression Inventory; STAI, State–Trait Anxiety Inventory (Trait Version); PSWQ, Penn State Worry Questionnaire; MIA, Mobility Inventory for Agoraphobia; SPS, Social Phobia Scale; PDSS, Panic Disorder Severity Scale; S.D., standard deviation; FU, follow-up; ES, effect size; tx, treatment.

[a] The *n* values shown reflect the number of cases available for analysis that included row symptom measure at both the baseline and the 4-month assessment.

[b] Change ES is calculated as the difference in the 4-month and baseline mean divided by the pooled S.D. (across CBT and PMRT) of the respective symptom measure at baseline.

[c] ANCOVA was performed controlling for baseline symptom measures. β is the regression coefficient for treatment (CBT *v.* PMRT) and indicates the augmentation in decrease of symptoms (on the original scales) found for CBT beyond PMRT and the *p* value is for the associated test of β equal to zero indicating no significant augmentation. The tx ES is calculated by taking the β and dividing by the pooled S.D. (across CBT and PMRT) of the respective symptom measure at baseline.

treatment effects on latent Distress. A practical implication of this finding is that if we were to pick a single indicator of distress to monitor closely throughout treatment, Trait Anxiety would provide the most information. These findings and their implications would not have been available in a traditional analysis of the manifest variables alone.

Our results also provide evidence that supports the validity of the basic empirical structure of internalizing psychopathology reported in earlier work. For example, we found the same basic empirical structure for continuous measures of internalizing symptom domains in clinical patients as that reported by Krueger & Markon (2006) in their review of community-based psychiatric epidemiology studies that factor analyze DSM diagnoses. We further extended this past work by showing that the same empirical structure identified at the baseline assessment when patients were highly symptomatic was also evident at follow-up when internalizing symptoms had decreased substantially. This is important because the question of whether the inter-relationship of measures (i.e. their meaning relative to one another) changes after an intervention has rarely been examined in the clinical literature; although it has be examined more extensively outside of the clinical literature (e.g. Millsap & Hartog, 1988).

Finding that the empirical structure of internalizing psychopathology is relatively insensitive to symptom intensity (i.e. before *versus* after treatment) or to the scope of measurement (i.e. symptom scale scores *versus* diagnoses) is consistent with the view that the latent variables identified are causal traits with regard to the observed indicators of internalizing psychopathology. With that said, it is important to consider the inherent uncertainty as to the causal status of latent constructs (e.g. Borsboom *et al*. 2003). Empirically, these latent constructs reflect the covariation structure of the manifest indicators. An assumption of our work, stemming from the psychiatric epidemiological work on which it was based (e.g. Krueger, 1999), is that these empirically identified factors can be interpreted such that status on the trait determines the probability of endorsing the symptoms assessed in the manifest measures in a particular way; however, continued research on the causal status of diagnostic rubrics in psychiatry, such as the Distress and Fear variables studied here, would be welcome (Krueger & Goldman, 2010).

It is also important to keep in mind that the CBT treatment effect on the Distress variable, although significant, was small and should be understood relative to the much larger decrease in Distress and Fear noted for both groups from baseline to follow-up. Regarding the decrease in internalizing symptoms experienced in both groups, we cannot partition that which was due to the study treatments from that which was due to the AUD treatment itself. For example, Brown *et al*. (1991) showed that anxiety and depression symptoms do decrease significantly in the weeks and months following initiation of AUD treatment; however, AUD treatment effects could not have accounted for the group difference on Distress. Additionally, analyses reported in this study were not intended to disambiguate the relationship between internalizing symptoms and alcohol use. For example, internalizing symptoms reported at baseline could have been the result of, the cause of, or unrelated to pathological alcohol use occurring prior to the study. Similarly, some participants had relapsed to drinking by the 4-month assessment and some had not, which could also have served as a cause or effect of internalizing symptoms measured at the follow-up.

To conclude, the broad aim of this work was to establish an analytic approach that addresses a broad, transdiagnostic spectrum of internalizing psychopathology in a clinical trial. Clinically, we assembled a CBT-based program to address internalizing psychopathology domains identified in earlier research as manifest in an AUD treatment patient sample. This is consistent with a growing movement embracing 'transdiagnostic' treatments for internalizing disorders (e.g. Barlow *et al*. 2004; Mansell *et al*. 2008, 2009; Ellard *et al*. 2010). Analytically, our approach was to harness the covariation that exists among related internalizing disorder symptom domains by establishing the best-fitting latent structure describing this covariation, and then quantifying each individual in terms of their level on the resulting latent variables both before and after the treatment. In this approach, correlated disorders are a non-issue because the empirical method is designed to partition overlapping variance without redundancy from any remaining unique variance (i.e. symptom elements that do not covary with other elements of the internalizing symptoms measured). This latent variable modeling approach is highly generalizable because it is applicable to any case in which observed variables are related to latent variables, regardless of the scope of measurement (e.g. symptoms, scale scores, diagnoses).

Because clinical trials provide a vector for transmitting taxonomic/measurement issues to clinical and policy decision making that affect patients' lives, we hope this work provides a pathway for considering how to optimally model internalizing psychopathology in clinical trials.

Acknowledgments

This work was supported by National Institute on Alcohol Abuse and Alcoholism (NIAAA) grants

R01-AA015069 and K02-AA0017886 awarded to M.G.K.

Declaration of Interest

None.

References

Andrews G, Goldberg DP, Krueger RF, Carpenter WT, Hyman SE, Sachdev P, Pine DS (2009). Exploring the feasibility of a meta-structure for DSM-V and ICD-11: could it improve utility and validity? *Psychological Medicine* **39**, 1993–2000.

Andrews G, Slade T, Issakidis C (2002). Deconstructing current comorbidity: data from the Australian National Survey of Mental Health and Well-Being. *British Journal of Psychiatry* **181**, 306–314.

APA (1980). *Diagnostic and Statistical Manual of Mental Disorders*, 3rd edn. American Psychiatric Association: Washington, DC.

Barlow DH, Allen LB, Choate ML (2004). Toward a unified treatment for emotional disorders. *Behavior Therapy* **35**, 205–230.

Barlow DH, Craske MG, Cerny JA, Klosko JS (1989). Behavioral treatment of panic disorder. *Behavior Therapy* **20**, 261–282.

Beck AT (1976). *Cognitive Therapy and the Emotional Disorders*. International Universities Press: New York.

Beck AT, Rush AJ, Shaw BF, Emery G (1979). *Cognitive Therapy of Depression*. Guilford Press: New York.

Beck AT, Ward CH, Mendelson M, Mock J, Erbaugh J (1961). An inventory for measuring depression. *Archives of General Psychiatry* **4**, 53–63.

Beekman AT, de Beurs E, van Balkom AJ, Deeg DJ, van Dyck R, van Tilburg W (2000). Anxiety and depression in later life: co-occurrence and communality of risk factors. *American Journal of Psychiatry* **157**, 89–95.

Bernstein DA, Borkovec TD (1973). *Progressive Relaxation Training. A Manual for the Helping Professions*. Research Press: Champaign, IL.

Borsboom D, Mellenbergh G, van Heerden J (2003). The theoretical status of latent variables. *Psychological Review* **110**, 203–219.

Brown EJ, Turovsky J, Heimberg RG, Juster HR, Brown TA, Barlow DH (1997). Validation of the Social Interaction Anxiety Scale and the Social Phobia Scale across the anxiety disorders. *Psychological Assessment* **9**, 21–27.

Brown SA, Irwin M, Schuckit MA (1991). Changes in anxiety among abstinent male alcoholics. *Journal of Studies on Alcohol* **52**, 55–61.

Brown TA, Campbell LA, Lehman CL, Grisham JR, Mancill RB (2001). Current and lifetime comorbidity of the DSM-IV anxiety and mood disorders in a large clinical sample. *Journal of Abnormal Psychology* **110**, 585–599.

Chambless DL, Caputo GC, Jasin SE, Gracely EJ, Williams C (1985). The Mobility Inventory for Agoraphobia. *Behaviour Research and Therapy* **23**, 35–44.

Eaton NR, Krueger RF, Oltmanns TF (2011). Aging and the structure and the long-term stability of the internalizing spectrum of personality and psychopathology. *Psychology and Aging* **26**, 987–993.

Ellard KK, Fairholme CP, Boisseau CL, Farchione TJ, Barlow DH (2010). Unified protocol for the transdiagnostic treatment of emotional disorders: protocol development and initial outcome data. *Cognitive Behavioral Practice* **17**, 88–101.

Ellis A, Harper RO (1975). *A New Guide to Rational Living*. Prentice-Hall: Oxford, UK.

Foa EB, Kozak MJ (1986). Emotional processing of fear: exposure to corrective information. *Psychological Bulletin* **99**, 20–35.

First MB, Spitzer R, Gibbon M, Williams J (1989). *Structural Clinical Interview for Axis-I DSM-IV Disorders Patient Edition (SCID-I/P Version 2.0)*. New York State Psychiatric Institution: New York, NY.

Heimberg RG, Dodge CS, Hope DA, Kennedy CR, Zollo LJ, Becker RE (1990). Cognitive behavioral group treatment for social phobia: comparison with a credible placebo control. *Cognitive Therapy and Research* **14**, 1–23.

Houck PR, Spiegel DA, Shear MK, Rucci P (2002). Reliability of the self-report version of the Panic Disorder Severity Scale. *Depression and Anxiety* **15**, 183–185.

Hu L, Bentler PM (1999). Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Structural Equation Modeling* **6**, 1–55.

Jacobson E (1938). *Progressive Relaxation*, 2nd edn. University of Chicago Press: Oxford, UK.

Kendler KS, Prescott CA, Myers J, Neale MC (2003). The structure of genetic and environmental risk factors for common psychiatric and substance use disorders in men and women. *Archives of General Psychiatry* **60**, 929–937.

Kessler RC, Chiu W, Demler O, Walters E (2005). Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the National Comorbidity Survey Replication. *Archives of General Psychiatry* **62**, 617–627.

Kessler RC, Ormel J, Petukhova M, McLaughlin KA, Green JG, Russo LJ, Stein DJ, Zaslavsky AM, Aguilar-Gaxiola S, Alonso J, Andrade L, Benjet C, de Girolamo G, de Graaf R, Demyttenaere K, Fayyad J, Haro JM, Hu CY, Karam A, Lee S, Lepine J, Matchsinger H, Mihaescu-Pintia C, Posada-Villa J, Sagar R, Üstün TB (2011). Development of lifetime comorbidity in the World Health Organization World Mental Health Surveys. *Archives of General Psychiatry* **68**, 90–100.

Kline RB (2010). *Principles and Practice of Structural Equation Modeling*. Guilford Press: New York.

Krueger RF (1999). The structure of common mental disorders. *Archives of General Psychiatry* **56**, 921–926.

Krueger RF, Finger MS (2001). Using item response theory to understand comorbidity among anxiety and unipolar mood disorders. *Psychological Assessment* **13**, 140–151.

Krueger RF, Goldman D (2010). The need for dimensional approaches in discerning the origins of psychopathology. In *Causality and Psychopathology: Finding the Determinants of Disorders and their Cures* (ed. P. Shrout, K. Keyes and K. Ornstein), pp. 338–351. Oxford University Press: New York.

Krueger RF, Markon KE (2006). Reinterpreting comorbidity: a model-based approach to understanding and classifying psychopathology. *Annual Review of Clinical Psychology* **2**, 111–133.

Kushner MG, Abrams K, Thuras P, Hanson KL, Brekke M, Sletten S (2005). Follow-up study of anxiety disorder and alcohol dependence in comorbid alcoholism treatment patients. *Alcoholism: Clinical and Experimental Research* **29**, 1432–1443.

Kushner MG, Maurer EW, Thuras P, Donahue C, Frye B, Menary KR, Hobbs J, Haeny AM, Van Demark J (in press). Hybrid cognitive-behavioral therapy versus relaxation training for co-occurring anxiety and alcohol disorder: a randomized clinical trial. *Journal of Consulting and Clinical Psychology*.

Kushner MG, Wall MM, Krueger RF, Sher KJ, Maurer E, Thuras P, Lee S (2012). Alcohol dependence is related to overall internalizing psychopathology load rather than to particular internalizing disorders: evidence from a national sample. *Alcoholism: Clinical and Experimental Research* **36**, 325–331.

Little RJA, Rubin DB (2002). *Statistical Analysis with Missing Data*, 2nd edn. John Wiley & Sons: New York.

Mansell W, Harvey A, Watkins ER, Shafran R (2008). Cognitive behavioral processes across psychological disorders: a review of the utility and validity of the transdiagnostic approach. *International Journal of Cognitive Therapy* **1**, 181–191.

Mansell W, Harvey A, Watkins E, Shafran R (2009). Conceptual foundations of the transdiagnostic approach to CBT. *Journal of Cognitive Psychotherapy* **23**, 6–19.

Markon KE (2010). Modeling psychopathology structure: a symptom-level analysis of Axis I and II disorders. *Psychological Medicine* **40**, 273–288.

Marsh HW, Muthén B, Asparouhov T, Lüdtke O, Robitzsch A, Morin AJS, Trautwein U (2009). Exploratory structural equation modeling, integrating CFA and EFA: application to students' evaluations of university teaching. *Structural Equation Modeling* **16**, 439–476.

McGlinchey JB, Zimmerman M (2007). Examining a dimensional representation of depression and anxiety disorders' comorbidity in psychiatric outpatients with item response modeling. *Journal of Abnormal Psychology* **116**, 464–474.

Meredith W (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika* **58**, 525–543.

Meyer TJ, Miller ML, Metzger RL, Borkovec TD (1990). Development and validation of the Penn State Worry Questionnaire. *Behaviour Research and Therapy* **28**, 487–495.

Millsap RE, Hartog SB (1988). Alpha, beta and gamma change in evaluation research: a structural equation approach. *Journal of Applied Psychology* **73**, 574–584.

Mineka S, Watson D, Clark LA (1998). Comorbidity of anxiety and unipolar mood disorders. *Annual Review of Psychology* **49**, 377–412.

Moffitt TE, Harrington H, Caspi A, Kim-Cohen J, Goldberg D, Gregory AM, Poulton R (2007). Depression and generalized anxiety disorder: cumulative and sequential comorbidity in a birth cohort followed prospectively to age 32 years. *Archives of General Psychiatry* **64**, 651–660.

Muthén LK, Muthén BO (2011). *Mplus User's Guide*, 6th edn. Muthén & Muthén: Los Angeles, CA.

Simms LJ, Prisciandaro JJ, Kruger RF, Goldberg DP (2012). The structure of depression, anxiety and somatic symptoms in primary care. *Psychological Medicine* **42**, 15–28.

Spielberger CD, Sydeman SJ, Owen AE, Marsh BJ (1999). Measuring anxiety and anger with the State-Trait Anxiety Inventory (STAI) and State-Trait Anger Expression Inventory (STAXI). In *The Use of Psychological Testing for Treatment Planning and Outcome Assessment*, 2nd edn (ed. M. E. Maruish), pp. 295–300. Lawrence Erlbaum Associates: Mahwah, NJ.

Vrieze SI (2012). Model selection and psychological theory: a discussion of the differences between the Akaike information criterion (AIC) and the Bayesian information criterion (BIC). *Psychological Methods* **17**, 228–243.

Watson D (2009). Differentiating the mood and anxiety disorders: a quadripartite model. *Annual Review of Clinical Psychology* **5**, 221–247.