Psychological Medicine (2009), **39**, 1087–1096. © 2008 Cambridge University Press
doi:10.1017/S0033291708004522 Printed in the United Kingdom

ORIGINAL ARTICLE

A comparison of the predictive abilities of dimensional and categorical models of unipolar depression in the National Comorbidity Survey

J. J. Prisciandaro and J. E. Roberts*

University at Buffalo, The State University of New York, Amherst, NY, USA

Background. Taxometric research on depression has yielded mixed results, with some studies supporting dimensional solutions and others supporting taxonic solutions. Although supplementary tests of construct validity might clarify these mixed findings, to date such analyses have not been reported. The present study represents a follow-up to our previous taxometric study of depression designed to evaluate the relative predictive validities of dimensional and categorical models of depression.

Method. Two sets of dimensional and categorical models of depression were constructed from the depression items of the Composite International Diagnostic Interview: (1) empirically derived models obtained using latent structure analyses and (2) rationally selected models, including an additive depressive symptoms scale (dimensional) and DSM major depressive episodes (categorical). Both sets of dimensional and categorical models were compared in terms of their abilities to predict various clinically relevant outcomes (psychiatric diagnoses and impairment).

Results. Factor analyses suggested a two-factor model ('cognitive–affective' and 'somatic' symptoms) and latent class analyses suggested a three-class model ('severe depression', 'moderate depression' and 'cognitive–affective distress'). In predictive analyses that simultaneously included dimensional and categorical models as predictors, the dimensional models remained significant unique predictors of outcomes while the categorical models did not.

Conclusions. Both dimensional models provided superior predictive validity relative to their categorical counterparts. These results provide construct validity evidence for the dimensional findings from our previous taxometric study and thus inspire confidence in dimensional conceptualizations of depression. It remains for future research to evaluate the construct validity of the taxonic solutions reported in the literature.

Received 7 August 2007; Revised 11 August 2008; Accepted 16 August 2008; First published online 10 October 2008

Key words: Categorical, depression, dimensional, National Comorbidity Survey, taxometric.

Introduction

For the past quarter-century, there has been considerable debate concerning whether depression is a dimensional or categorical construct. Proponents of a dimensional perspective suggest that 'clinical' depression represents elevation on continua of depressive symptom severity (e.g. Flett *et al*. 1997). Proponents of a categorical perspective maintain that clinical depression is discontinuous from normal functioning and subsyndromal conditions (e.g. Coyne, 1994). This debate has been fueled in part by the importance of the latent structure of depression for research and assessment. Although the prevailing psychiatric diagnostic system (DSM-IV; APA, 1994) has taken a categorical perspective, the preponderance of empirical research has supported a dimensional model. In addition to research demonstrating the continuity of depressive symptoms and correlates among individuals varying in their diagnostic classification or degree of depressive severity (reviewed in Flett *et al*. 1997), taxometric methods (Waller & Meehl, 1998) have been employed to directly test whether depression is categorical or dimensional.

A number of taxometric studies have suggested that there are no discontinuities in the latent structure of depression (e.g. Ruscio & Ruscio, 2000, 2002; Franklin *et al*. 2002; Hankin *et al*. 2005; Prisciandaro & Roberts, 2005; Slade & Andrews, 2005). For example, Prisciandaro & Roberts (2005) conducted a taxometric analysis of unipolar depression using structured diagnostic interview data (Composite International Diagnostic Interview, CIDI; Robins *et al*. 1988) in a large community sample (National Comorbidity Survey, NCS; Kessler, 2002*a*) and found that depression is best

* Address for correspondence: Dr J. E. Roberts, University at Buffalo, The State University of New York, Department of Psychology, Park Hall, Box 604110, Amherst, NY 14260-4110, USA.
(Email: robertsj@buffalo.edu)

characterized by a dimensional structure. Alternatively, several taxometric studies have found evidence for categorical structures of depression (e.g. Solomon *et al*. 2006; Ruscio *et al*. 2007), depression-related constructs (e.g. 'Involuntary Defeat Syndrome'; Beach & Amir, 2006) and subtypes of depression (e.g. Haslam & Beck, 1994; Ambrosini *et al*. 2002). Given these mixed findings, research is needed to corroborate previous taxometric findings using traditional methods of construct validation to ensure that results were not artifactual (e.g. determined by distributional properties of the data; Watson, 2003).

One approach to evaluating the validity of past taxometric studies would be to compare the relative abilities of dimensional and categorical models of depression to predict theoretically related constructs. If depression is a dimensional construct, then dimensional models should demonstrate superior predictive validity relative to categorical models because of the loss of statistical power associated with dichotomizing a truly continuous variable (Cohen *et al*. 2002). However, if depression is a categorical construct, then categorical models should demonstrate superior ability to predict related constructs because the additional variability provided by the latter would reflect measurement error that reduces statistical power (Ruscio *et al*. 2006). Aggen *et al*. (2005) used a similar approach to evaluate the relative predictive abilities of a liability scale of depressive symptoms and diagnoses of depression. They found that diagnoses did not significantly predict neuroticism and future depressive episodes once the dimensional model was statistically accounted for, and that the dimensional model provided more precise estimates of parameters in twin models.

The present study was designed to examine the construct validity of the dimensional results in the Prisciandaro & Roberts (2005) taxometric study by evaluating the relative predictive abilities of categorical and dimensional models of depression in the NCS. In part I of the present study, categorical and dimensional models of depression were empirically derived from the NCS data. In part II, the empirically derived models (along with a pair of rationally selected categorical and dimensional models) were compared in terms of their abilities to predict variance in clinically relevant constructs.

Part I – Method

Sample and measure

The NCS (Kessler *et al*. 1994) obtained a stratified probability sample of the United States population ($n=8098$) and administered a structured diagnostic interview, the CIDI (Robins *et al*. 1988), which included an assessment of participants' depressive symptoms according to DSM-III-R criteria (APA, 1987). Inter-rater reliability of depressive symptoms in the CIDI has been found to be satisfactory (κ's ranged from 0.69 to >0.90; Wacker *et al*. 1990; Wittchen, 1991). Test–retest and inter-rater reliabilities of diagnoses of depressive disorders were also good ($\kappa=0.71$ and 0.95, respectively; Wittchen, 1994). Concordance between CIDI and Structured Clinical Interview for DSM-III-R (Spitzer *et al*. 1992) diagnoses of major depressive episodes (MDEs) were acceptable ($\kappa=0.53$; Kessler *et al*. 1998).

Indicators of unipolar depression were drawn from the section of the CIDI that asked participants to report symptoms of depression from a 2-week period of their lives when they experienced some degree of symptomatology. A subsample of the NCS was used for the present study because only participants who reported the lifetime experience of three or more co-occurring depressive symptoms (including either sad mood or anhedonia) were administered this section of the CIDI. The binary depressive symptom items from the CIDI were aggregated into the nine DSM diagnostic symptoms of an MDE [depressed mood, diminished interest or pleasure ('anhedonia'), appetite or weight loss or gain ('weight disturbance'), insomnia or hypersomnia ('sleep disturbance'), psychomotor agitation or retardation ('psychomotor disturbance'), fatigue or loss of energy ('fatigue'), feelings of worthlessness or guilt ('worthlessness/guilt'), impaired concentration or indecisiveness ('concentration/indecisiveness'), recurrent thoughts of death or suicide, suicide plan, or attempt ('suicidality'); APA, 1987] according to the NCS diagnostic algorithm (Zhao *et al*. 1994). This algorithm required that affirmatively coded symptoms were not 'due entirely to medications, drugs, alcohol, physical illness or injury'. The final sample ($n=2803$) contained more females than males (59.1% *v*. 40.1%), was predominantly White (79.8%; 8.8% Black, 8.5% Hispanic and 2.9% other), ranged in age from 15 to 58 years, with a mean age of 33.9 years (S.D. = 10.2 years), and had a 51.7% diagnostic base rate of MDEs.

Analytic strategy

Latent structure analyses were conducted to empirically derive categorical (via latent class analysis) and dimensional (via factor analysis) models from the data. All analyses incorporated appropriate NCS sample weights (Kessler *et al*. 1994). The selected sample was randomly divided into two subsamples ($n_1=1402$, $n_2=1401$). Missing data were negligible ($<2\%$) and were list-wise deleted. Structural analyses

were conducted using Mplus version 5.1 (Muthén & Muthén, 2007).

Factor analyses

Exploratory factor analysis (EFA) models were estimated in the first subsample from the tetrachoric correlation matrix of the nine dichotomous depressive symptom items using robust weighted least-squares estimation (WLSMV; Muthén, 2004). Parallel analysis (Horn, 1965) was conducted using 1000 sets of random data (O'Connor, 2000). Factor models were rotated using a promax (oblique) rotation method (Preacher & MacCallum, 2003). The obtained EFA model was used to construct a confirmatory factor analysis (CFA) model in the second subsample using WLSMV estimation. Each symptom item was loaded on the factor on which it had the highest loading in the EFA; cross-loadings were included if secondary loadings in the EFA exceeded 0.32 (Comrey & Lee, 1992), and factors were free to co-vary. CFA analyses incorporated information regarding the stratification and clustering of the NCS data (Muthén, 2004). Model fit was evaluated according to established standards [comparative fit index (CFI) >0.95, Tucker–Lewis index (TLI) >0.95, root mean square error of approximation (RMSEA) <0.05; Hu & Bentler, 1999]. An approximate χ^2 difference test (known as DIFFTEST; Muthén, 2004) was conducted to evaluate whether the final CFA model provided better fit to the data than a more parsimonious one-factor model.

Latent class analyses

Latent class analysis (LCA) models were estimated in the first subsample from the nine depressive symptom items using maximum likelihood estimation (Muthén, 2004). LCA analyses incorporated information regarding the stratification and clustering of the data (Muthén, 2004). LCA models with varying numbers of latent classes were estimated, and the model with the lowest Bayesian information criteria (BIC) was selected as the best-fitting model (Nylund *et al*. 2007). The magnitudes of differences in BIC between models were interpreted using descriptive guidelines (Raftery, 1993) that can be applied to non-nested models as rough rules of thumb (AE Raftery, personal communication, 7 December 2006). A 0- to 4.6-point difference provides 'weak' evidence in favor of the model with the smaller BIC; a 4.6- to 9.2-point difference provides 'strong' evidence; and a difference of greater than 9.2 provides 'conclusive' evidence. This procedure was repeated in the second subsample to evaluate the robustness of the best-fitting LCA model.

Part I – Results and Discussion

Factor analyses

Parallel analysis suggested that up to two factors should be extracted from the data [research data's eigenvalues = 3.426, 1.316, 1.070; random data's mean (95th percentile) eigenvalues = 1.123 (1.158), 1.083 (1.109), 1.052 (1.074)][1]†. Thus, a two-factor model was estimated. Loadings are presented in Table 1. The first factor was marked by 'cognitive–affective' items (e.g. depressed mood, worthlessness/guilt). The second factor was marked by 'somatic' items (e.g. weight disturbance, fatigue) and anhedonia. Factors were correlated ($r=0.40$) and the fit of the model was good (RMSEA = 0.03). The two-factor CFA model constructed from these findings provided acceptable fit to the second subsample data ($\chi^2=39.71$, df = 14, $p<0.001$, CFI = 0.94, TLI = 0.94, RMSEA = 0.02). All factor loadings were statistically significant (mean loading = 0.59, mean $R^2=0.37$) (See Table 1). Factors were correlated ($r=0.52$). Finally, the two-factor model provided a superior fit to the data than a unidimensional model ($\Delta\chi^2=29.21$, df = 1, $p<0.001$)[2].

Latent class analyses

Comparison of BIC values from two- (BIC = 14 645.38), three- (BIC = 14 585.25) and four- (BIC = 14 625.53) class LCA models provided 'conclusive' evidence for the three-class model's superior relative fit in the first subsample of data. Item response probabilities are presented in Table 1. Class 1 ($n=362$, 'severe depression') consisted of individuals with uniformly high response probabilities across all symptoms (mean probability = 0.84) and class 2 ($n=719$, 'moderate depression') consisted of individuals with moderate response probabilities (mean probability = 0.48). Of individuals assigned to the 'severe depression' class, 98% were diagnosed with an MDE, in contrast to 44% of those assigned to 'moderate depression'. Class 3 ($n=321$, 'cognitive–affective distress') consisted of individuals with higher response probabilities for the 'cognitive–affective' symptoms of depression than class 2 (mean probability = 0.63 *v.* 0.53) but with uniformly low response probabilities for the 'somatic' symptoms (mean probability = 0.13). Less than 1% of individuals assigned to the 'cognitive–affective distress' class were diagnosed with an MDE. Because item response probabilities appeared roughly ordered across classes, a three-class discrete metrical model (Markon & Krueger, 2006) was estimated, with classes ordered in terms of depressive severity (Croon, 2002). Despite the apparent ordering of response

† The notes appear on p. 1095.

Table 1. *EFA and CFA factor loadings and R^2 values, and LCA item response probabilities for the nine DSM criteria of a major depressive episode*[a]

	EFA		CFA			LCA		
	Factor 1	Factor 2	Factor 1	Factor 2		Class 1	Class 2	Class 3
Indicator	'Cognitive–affective'	'Somatic'	'Cognitive–affective'	'Somatic'	R^2	'Severe'	'Moderate'	'Distress'
Worthlessness/guilt	0.56[b]	0.07	0.74	–	0.55	0.73	0.23	0.43
Suicidality	0.47[b]	0.07	0.53	–	0.28	0.74	0.45	0.55
Depressed mood	0.38[b]	−0.09	0.23	–	0.05	0.99	0.90	1.00
Weight disturbance	−0.16	0.69[b]	–	0.54	0.29	0.81	0.54	0.07
Sleep disturbance	0.04	0.68[b]	–	0.63	0.40	0.97	0.66	0.26
Fatigue	0.01	0.67[b]	–	0.65	0.42	0.87	0.42	0.11
Psychomotor disturbance	0.23	0.61[b]	–	0.72	0.52	0.64	0.17	0.02
Concentration/indecisiveness	0.20	0.60[b]	–	0.66	0.44	0.95	0.43	0.17
Anhedonia	0.09	0.53[b]	–	0.59	0.34	0.82	0.48	0.16

EFA, Exploratory factor analysis; CFA, confirmatory factor analysis; LCA, latent class analysis; Severe, severe depression; Moderate, moderate depression; Distress, cognitive–affective distress.

[a] EFA and LCA results are from subsample 1 of the data ($n = 1402$). CFA results are from subsample 2 ($n = 1401$).

[b] EFA factor loadings > 0.32.

probabilities, the nominal three-class LCA model (BIC = 14 585.25) demonstrated 'conclusive' evidence for superior fit to the three-class discrete metrical model (BIC = 14 605.00). In the second subsample, the three-class (BIC = 14 811.81) LCA model again demonstrated conclusive evidence for superior fit to the two- (BIC = 14 913.01) and four- (BIC = 14 837.14) class models. Furthermore, the three-class LCA model (BIC = 14 811.81) again demonstrated 'conclusive' evidence for superior fit to the three-class discrete metrical model (BIC = 14 890.76).

Part II – Introduction

Part II of the present study investigated the relative abilities of dimensional and categorical models of depression to predict variance in clinically relevant constructs. Two sets of model comparisons were made across outcomes: (1) Between the empirically derived models from part I; and (2) between rationally selected models. For a dimensional model, we created a single additive scale with each depressive symptom contributing 1 point. For a categorical model, we chose the pre-eminent diagnostic model of depression: DSM major depressive disorder (MDD).

Two types of outcomes[3] were selected: psychiatric diagnoses and impairment. Psychiatric diagnoses were selected because approximately 75% of individuals with lifetime MDD meet diagnostic criteria for at least one other psychiatric disorder (Kessler *et al*. 2003) and because the NCS was expressly designed to accurately estimate the prevalence of psychiatric diagnoses. Impairment outcomes were selected because MDD is a pervasively impairing disorder associated with divorce (Wade & Cairney, 2000), college dropout (Kessler *et al*. 1995), health problems (Sullivan *et al*. 2001) and suicide (e.g. Isomestä *et al*. 1994).

The main criterion used to evaluate the models' relative predictive abilities was whether the categorical model significantly predicted unique variance in a given outcome when the dimensional predictor was simultaneously considered (and vice versa).

Part II – Method

Sample and measure

Analyses involving outcomes from the first stage of the NCS used the selected sample described in part I. Analyses involving second-stage NCS variables used a further reduced subsample (maximum $n = 2609$) (Kessler *et al*. 1997). An additional NCS sample weight was used to adjust for participants' unequal probability of being selected into the second stage.

The empirically derived dimensional model from part I was represented as two continuous variables in part II by computing factor scores from the final

Table 2. *Multiple regression analyses with empirically derived dimensional or categorical models of depression alone predicting four outcomes, followed by analyses with both models of depression simultaneously predicting outcomes*[a]

	Role impairment		Treatment-seeking		Internalizing		Externalizing	
	ΔR^2	*B*	ΔR^2	*B*	ΔR^2	*B*	ΔR^2	*B*
I. Addition of categorical predictors to models with dimensional predictors only								
Step 1								
Cognitive–emotional dimensional	0.13*	0.37*	0.06*	0.31*	0.10*	0.22*	0.06*	0.38*
Somatic dimensional		0.07		−0.05		0.01		−0.17*
Step 2								
Cognitive–emotional dimensional	0.01*	0.30*	0.00	0.28*	0.00	0.20*	0.00	0.34*
Somatic dimensional		0.15		−0.03		−0.02		−0.12
Moderate depression categorical		−0.04		−0.03		−0.07		−0.02
Cognitive–emotional categorical		0.11		0.02		−0.08		0.06
II. Addition of dimensional predictors to models with categorical predictors only								
Step 1								
Moderate depression categorical	0.09*	−0.37*	0.03*	−0.19*	0.07*	−0.19*	0.02*	−0.15*
Cognitive–emotional categorical		−0.34*		−0.15*		−0.20*		−0.01
Step 2								
Moderate depression categorical	0.05*	−0.04	0.03*	−0.03	0.04*	−0.07	0.04*	−0.02
Cognitive–emotional categorical		0.11		0.02		−0.08		0.06
Cognitive–emotional dimensional		0.30*		0.28*		0.20*		0.34*
Somatic dimensional		0.15		−0.03		−0.02		−0.12

Internalizing, Internalizing disorders; Externalizing, externalizing disorders; ΔR^2, change in proportion of variance accounted for; *B*, unstandardized regression coefficient.

[a] Categorical predictors were dummy coded such that the 'severe depression' class was the reference group ('0') for each contrast.

* $p < 0.001$.

two-factor CFA model. The three-class LCA model in part I was represented as two binary contrast variables in part II by computing participants' category membership from the final LCA model, and by subsequently dummy-coding category membership with the 'severe depression' class as the reference. 'Severe depression' was chosen as the reference class because it was the most well known of the classes, as 98% met criteria for MDE. The rationally selected dimensional model was a single additive scale with each depressive symptom contributing 1 point. The rationally selected categorical model of depression was lifetime DSM-III-R diagnoses of MDEs derived from the NCS diagnostic algorithm based on the same 2-week symptom data used for the dimensional models.

Outcome variables

Psychiatric diagnoses were assigned on a lifetime basis and were coded '0' = absent, '1' = present. DSM-III-R hierarchy rules were observed. To reduce diagnostic outcomes, Krueger's (1999) dimensional model of common mental illnesses was estimated using CFA, and the resulting two factors (internalizing and externalizing disorders) were used as outcomes in the present study. Agoraphobia, generalized anxiety disorder, panic disorder, simple phobia and social phobia were loaded on the 'internalizing' factor; alcohol dependence, antisocial personality disorder and drug dependence were loaded on the 'externalizing factor'. Factors were free to co-vary.

Role impairment-related variables included: 'interference with activities' ['How much did your period(s) of feeling (sad and/or uninterested) ever interfere with your life or activities?' ('1', not at all to '4', a lot)]; and 'social and work impairment' ['Was any period of feeling (sad and/or uninterested) so bad that it kept you from working or from seeing friends or relatives?' ('0' = no, '1' = yes)]. Treatment-seeking variables included: 'mental health professional' ['Did you ever … see (a psychiatrist, psychologist, social worker, or counselor) for problems with your emotions or nerves or your use of alcohol or drugs?' ('0' = no, '1' = yes)]; 'psychiatric hospitalization' [lifetime overnight admission to a hospital or admission to a hospital emergency department for help with 'emotions or nerves' or 'use of alcohol or drugs' ('0' = no, '1' = yes)]; and 'psychiatric medications' ['In the past 12 months,

Table 3. *Multiple regression analyses with rationally selected dimensional or categorical models of depression alone predicting four outcomes, followed by analyses with both models of depression simultaneously predicting outcomes*

	Role impairment		Treatment-seeking		Internalizing		Externalizing	
	ΔR^2	*B*	ΔR^2	*B*	ΔR^2	*B*	ΔR^2	*B*
I. Addition of categorical predictors to models with dimensional predictors only								
Step 1								
Depressive symptoms scale	0.11*	0.08*	0.03*	0.04*	0.08*	0.04*	0.01	0.02
Step 2								
Depressive symptoms scale	0.00	0.09*	0.01*	0.06*	0.02*	0.06*	0.01*	0.04*
Major depressive episodes		−0.04		−0.11*		−0.12*		−0.11*
II. Addition of dimensional predictors to models with categorical predictors only								
Step 1								
Major depressive episodes	0.05*	0.23*	0.00	0.07	0.01	0.06	0.00	0.01
Step 2								
Major depressive episodes	0.08*	−0.04	0.04*	−0.11*	0.10*	−0.12*	0.02*	−0.11*
Depressive symptoms scale		0.09*		0.06*		0.06*		0.04*

Internalizing, Internalizing disorders; Externalizing, externalizing disorders; ΔR^2, change in proportion of variance accounted for; *B*, unstandardized regression coefficient.

* $p < 0.001$.

did you take any of the following …: sedatives, anti-depressants, tranquilizers, stimulants, painkillers, or anti-psychotics; under the supervision of a doctor, for your emotions or nerves or your use of alcohol or drugs?' ('0' = no, '1' = yes)]. Two latent factors were estimated from these variables using CFA: (1) 'role impairment' and (2) 'treatment-seeking'. The resulting factors were used as outcomes.

Analytic strategy

Factor analyses

As described above, Krueger's (1999) two-factor model was estimated for diagnostic outcomes using CFA, whereas separate unidimensional models were estimated for 'role impairment' and 'treatment-seeking' outcomes. CFA analyses incorporated the NCS sample weight as well as information regarding the stratification and clustering of the data (Muthén, 2004). Model fit was evaluated according to established standards (CFI > 0.95, TLI > 0.95, RMSEA < 0.05; Hu & Bentler, 1999). Outcomes were derived from the above CFA models by computing factor scores from each model for all participants.

Regression analyses

Regression models were estimated using the %SREGSUB macro (SAS Institute, Inc., 2002) in SAS version 9.1.3 (SAS Institute Inc., Cary, NC, USA), which allows for the analysis of a subpopulation of complex survey data and includes information regarding the weighting, stratification and clustering of the data. Separate set-wise regression models were conducted for each of the four outcome variables (role impairment, treatment-seeking, internalizing and externalizing). Step 1 included either a set of dimensional variables or a set of categorical variables, whereas step 2 added the set of variables not represented in step 1 (e.g. the categorical variables would be added at step 2 if the dimensional variables had been included at step 1). The improvement in model fit at step 2 (represented by ΔR^2 and assessed with an *F* test) represents the degree to which one model of depression (e.g. categorical) provided unique predictive validity beyond that of the alternate model (e.g. dimensional). Additionally, it was of interest to determine which depression variables provided unique contributions to the prediction of outcome variables as reflected in the β values at step 2. Separate Bonferroni corrections were applied to significance tests of ΔR^2 and β's to control experiment-wise α inflation. With a desired α of 0.05 for each type of test, Bonferroni corrections suggested an α of 0.002 (0.05/32) for ΔR^2 tests and an α of 0.001 (0.05/48) for tests of β values. However, all results are reported at an α of 0.001 for simplicity given that all ΔR^2 tests that were significant at $\alpha = 0.002$ were also significant at $\alpha = 0.001$. Missing data were negligible (< 2%) and were list-wise deleted.

Part II – Results and Discussion

Factor analyses

Krueger's (1999) model of mental illnesses provided good fit to the data ($\chi^2 = 25.92$, df = 9, $p < 0.01$, CFI = 0.96, TLI = 0.96, RMSEA = 0.02). All factor loadings were statistically significant (mean loading = 0.69, mean $R^2 = 0.49$). The role impairment CFA model was just identified and therefore had perfect fit to the data. Factor loadings were statistically significant (mean loading = 0.75, mean $R^2 = 0.57$). Finally, the treatment-seeking CFA model also provided perfect fit to the data. Factor loadings were statistically significant (mean loading = 0.72, mean $R^2 = 0.52$).

Regression analyses

Empirically derived models

Results regarding the empirically derived models' abilities to predict psychiatric diagnoses are presented in Table 2. As can be seen in the top half of this table, the inclusion of the latent class contrasts at step 2 was not statistically significant in the case of treatment-seeking, internalizing or externalizing. In other words, the categorical variables did not lead to a statistically significant increase in R^2 compared with a model that only included dimensional predictors. Furthermore, none of the latent class contrasts provided statistically significant unique contributions to the prediction of any of the outcome variables after accounting for the dimensional variables. On the other hand, the inclusion of the latent class contrasts at step 2 made a statistically significant contribution to the prediction of role impairment. However, this contribution was very small in magnitude (R^2 change = 0.01) and neither of the latent class contrasts made a statistically significant unique contribution after controlling for the dimensional predictors.

As can be seen in the bottom half of Table 2, the inclusion of the latent dimensional variables at step 2 led to a statistically significant increase in R^2 in the case of all four outcome variables. Furthermore, the dimensional cognitive–affective symptoms variable made unique contributions to the prediction of all four outcome variables after controlling for the contributions of the latent class contrasts.

Rationally selected models

Results regarding rationally selected models' abilities to predict psychiatric diagnoses are presented in Table 3. As can be seen in the top half of this table, the inclusion of MDE at step 2 was statistically significant in the case of treatment-seeking, internalizing and externalizing. In other words, MDE made a statistically significant contribution to the prediction of these outcomes controlling for the dimensional symptom count. However, the negative β values indicate that MDEs were associated with less treatment-seeking, internalizing symptomatology and externalizing symptomatology adjusting for severity of depressive symptoms. These effects appear to be the result of statistical suppression as discussed below. In contrast, the inclusion of MDE at step 2 was not statistically significant in the case of role impairment.

As can be seen in the bottom half of Table 3, the inclusion of the dimensional symptom count at step 2 was statistically significant with each of the four outcome variables. Higher depressive symptomatology was associated with greater role impairment, treatment-seeking, internalizing symptomatology and externalizing symptomatology adjusting for the presence or absence of MDE.

In summary, the latent dimensional models of depression consistently demonstrated superior predictive validity relative to the latent categorical models.

Overall discussion

The results from the present study uniquely inform the continuity debate in the depression literature by testing the construct validity of a previous taxometric analysis (Prisciandaro & Roberts, 2005). Supplementary tests of construct validity can potentially clarify mixed findings in the taxometric literature, but have not previously been reported in the depression literature. The present study provides evidence that dimensional models of depression have greater predictive validity than the DSM's categorical model, as well as empirically derived latent class models, and suggests it is unlikely that the results from the Prisciandaro & Roberts (2005) study were artifactual. These results were consistent with those of another recent study that found that diagnoses of depression have little predictive validity after controlling for symptom severity in a sample of female twins (Aggen *et al*. 2005). It remains for future research to evaluate the construct validity of taxonic solutions that have been reported in the literature. For example, does the 'Involuntary Defeat Syndrome' (IDS) taxon (Beach & Amir, 2006) provide superior predictive ability compared with a dimensional representation of IDS symptoms? If not, it would be difficult to maintain that this construct has a latent categorical structure.

Although the DSM categorical variable predicted unique variance in several outcome variables after controlling for symptom severity, these effects appear to be due to statistical suppression. Specifically, MDEs predicted higher scores on internalizing symptoms,

externalizing symptoms and treatment-seeking when tested in the absence of the dimensional depressive symptomatology scale. In contrast, MDEs predicted lower scores on each of these variables when severity of symptoms was included in the model. In other words, after adjusting for number of symptoms, having a diagnosis of MDD was associated with less symptomatology and lower frequency of treatment seeking. Such sign reversals reflect statistical suppression (Cohen *et al*. 2002). One explanation for suppression in these data involves the hierarchy rules imposed by the DSM; it may be that all that remains of an MDE diagnosis after controlling for symptom severity are the DSM's exclusionary criteria. Specifically, individuals high in depressive symptoms with certain other more serious disorders (e.g. psychotic and bipolar disorders) would not have met criteria for MDE because these disorders were part of the exclusionary criteria for MDE. In contrast, individuals who met criteria for MDE by definition would not have had these other more impairing disorders.

Because of the structure of the CIDI, a major limitation of the present study was its reliance on a sample with elevated depressive symptoms, which prevents clear generalization of the obtained results to the broader population. For example, the factor analytic results from the present study may not reflect the dimensional structure of depression in the unselected population, making the relevance of the predictive ability of the obtained factor model to the community questionable. Nonetheless, the obtained model is highly congruous with the results of Shafer's (2006) meta-analysis of four widely used measures of depressive symptoms[4]. Specifically, the factors in the present study's two-factor model (i.e. 'cognitive–affective' and 'somatic' symptoms) closely resemble the two common factors that Shafer found across populations and across measures of depression: 'general depression' (e.g. guilt) and 'somatic symptoms' (e.g. insomnia). Additionally, analyses of the predictive ability of a more conservative unidimensional model of depression in the present study suggested that this dimensional model still provided superior predictive validity relative to the DSM model of depression. In addition to the above concern, LCA would not have been able to detect depressive taxa with extremely high base rates in the general population because we removed all of the potential complement members from the sample.

Related to the issues above, the present sample had a restricted range of depressive symptoms resulting from the removal of individuals with few depressive symptoms. However, the impact of this range restriction on our findings is predictable and arguably suggests that the superior predictive ability of the dimensional models would have been even greater in an unrestricted sample. Specifically, range restriction would have attenuated relationships between dimensional models and outcomes. In contrast, this restriction would have increased our ability to find support for categorical models because it raised the base rate of MDE to close to 50%, which should optimize its predictive ability (Cohen *et al*. 2002). Along the same vein, our final sample over-represented individuals with high levels of symptoms who nonetheless did not meet criteria for MDE. However, if the DSM's categorical model is valid, this oversampling should have had little impact on the ability of MDEs to predict theoretically relevant constructs; the DSM reflects a discrete categorical model that purposefully discards within-class dimensional variation on the grounds that individuals are relatively homogeneous within-class. This is especially true for the non-disordered class (which is the only class where range restriction occurred in the present study) because the DSM's symptom severity specifier only applies to individuals who met diagnostic criteria.

If future research supports the present study's findings, additional research should further examine the dimensional structure of depressive symptoms using factor analytic techniques in large representative samples. Alongside these efforts, assessment devices should be developed to discern individuals' proper placement on the latent dimensions of depressive symptomatology, as existing dimensional measures may not adequately reflect the latent structure of depression. A dimensional framework requires rethinking the allocation of treatment resources. One possibility involves setting a cut-point on the latent symptom dimensions based on functional impairment (Kessler, 2002*b*). If the relationship between depression and impairment is linear, cost–benefit analyses could determine the advantages and disadvantages of treating different levels of depressive severity. If the relationship between depressive severity and impairment is non-linear, then the inflection point of this relationship could be used as a practical diagnostic cut-point.

In summary, the present study demonstrated that dimensional models of depression provided superior predictive validity relative to categorical models. These results provide construct validity evidence for the dimensional findings from Prisciandaro & Roberts' (2005) taxometric study and thus inspire confidence in dimensional conceptualizations of depression. It remains for future research to evaluate the construct validity of taxonic solutions that have been reported in the literature.

Acknowledgements

The NCS is funded by the National Institute of Mental Health (grants R01 MH/DA46376 and R01 MH49098), the National Institute of Drug Abuse (through a supplement to R01 MH/DA46376) and the W. T. Grant Foundation (grant 90135190). We thank Craig Colder, Len Simms and John Ruscio for their helpful comments on an earlier draft of the paper. We are also grateful to Craig Colder, Bengt Muthén and Linda Muthén for their statistical advice.

Declaration of Interest

None.

Notes

[1] The percentages of variance in indicators explained by factors are not presented because the goal in exploratory factor analysis is to reproduce indicators' correlations as closely as possible, not to explain optimal amounts of variance (Preacher & MacCallum, 2003; and, thus, Mplus does not provide this information).

[2] The unidimensional CFA model did not fit well (CFI = 0.92, TLI = 0.91, RMSEA = 0.02) and the depressed mood item's loading (0.03) was non-significant.

[3] In the present study, the term 'outcome' refers to dependent variables in regression models. Although the tested models were necessarily stated in causal terms (e.g. a scale 'predicting an outcome'), causal relations between variables were not hypothesized. Models were consistently structured as depression 'predicting outcomes' because we were interested in examining the unique relationships among depression symptom scales and their ability to account for variance in other variables.

[4] The measures included in Shafer's (2006) meta-analysis were the Beck Depression Inventory, the Center for Epidemiologic Studies Depression Scale, the Hamilton Rating Scale for Depression and the Zung Self-Rating Depression Scale.

References

Aggen SH, Neale MC, Kendler KS (2005). DSM criteria for major depression: evaluating symptom patterns using latent-trait item response models. *Psychological Medicine* **35**, 475–487.

Ambrosini PJ, Bennett DS, Cleland CM, Haslam N (2002). Taxonicity of adolescent melancholia: a categorical or dimensional construct. *Journal of Psychiatric Research* **36**, 247–256.

APA (1987). *Diagnostic and Statistical Manual of Mental Disorders*, 3rd edn, revised. American Psychological Association: Washington, DC.

APA (1994). *Diagnostic and Statistical Manual of Mental Disorders*, 4th edn. American Psychological Association: Washington, DC.

Beach SRH, Amir N (2006). Depression is taxonic. *Journal of Psychopathology and Behavioral Assessment* **28**, 171–178.

Cohen J, Cohen P, West SG, Aiken LS (2002). *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*, 3rd edn. Erlbaum: Mahwah, New Jersey.

Comrey AL, Lee HB (1992). *A First Course in Factor Analysis*, 2nd edn. Erlbaum: Mahwah, New Jersey.

Coyne JC (1994). Self-reported distress: analog or ersatz depression? *Psychological Bulletin* **116**, 29–45.

Croon M (2002). Ordering the classes. In *Applied Latent Class Analysis* (ed. J. A. Hagenaars and A. L. McCutcheon), pp. 137–162. Cambridge University Press: Cambridge, UK.

Flett GL, Vredenburg K, Krames L (1997). The continuity of depression in clinical and nonclinical samples. *Psychological Bulletin* **121**, 395–416.

Franklin CL, Strong DR, Greene RL (2002). A taxometric analysis of the MMPI-2 depression scales. *Journal of Personality Assessment* **79**, 110–121.

Hankin BL, Fraley RC, Lahey BB, Waldman ID (2005). Is depression best viewed as a continuum or discrete category? A taxometric analysis of childhood and adolescent depression in a population-based sample. *Journal of Abnormal Psychology* **114**, 96–110.

Haslam N, Beck AT (1994). Subtyping major depression: a taxometric analysis. *Journal of Abnormal Psychology* **103**, 686–692.

Horn JL (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika* **30**, 179–185.

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.

Isometsä ET, Henriksson MM, Hillevi ME, Kuoppasalmi KI, Lönnqvist JK (1994). Suicide in major depression. *American Journal of Psychiatry* **151**, 530–536.

Kessler RC (2002*a*). *National Comorbidity Survey, 1990–1992* [*Data file*]. Conducted by University of Michigan, Survey Research Center, 2nd ICPSR edn. Inter-University Consortium for Political and Social Research: Ann Arbor, MI.

Kessler RC (2002*b*). Epidemiological perspectives for the development of future diagnostic systems. *Psychopathology* **35**, 158–161.

Kessler RC, Berglund P, Demler O, Jin R, Koretz D, Merikangas KR, Rush AJ, Walters EE, Wang PS (2003). The epidemiology of major depressive disorder: results from the National Comorbidity Survey Replication (NCS-R). *Journal of the American Medical Association* **289**, 3095–3105.

Kessler RC, Davis CG, Kendler KS (1997). Childhood adversity and adult psychiatric disorder in the US National Comorbidity Survey. *Psychological Medicine* **27**, 1101–1119.

Kessler RC, Foster CL, Saunders WB, Stang PE (1995). Social consequences of psychiatric disorders: I. Educational attainment. *American Journal of Psychiatry* **152**, 1026–1032.

Kessler RC, McGonagle KA, Zhao S, Nelson CB, Hughes M, Eshleman S, Wittchen HU, Kendler KS (1994). Lifetime and 12-month prevalence of DSM-III-R psychiatric disorders in the United States: results from the National Comorbidity Survey. *Archives of General Psychiatry* **51**, 8–19.

Kessler RC, Wittchen H-U, Abelson JM, McGonagle KA, Schwarz N, Kendler KS, Knäuper B, Zhao S (1998). Methodological studies of the Composite International Diagnostic Interview (CIDI) in the US National Comorbidity Survey (NCS). *International Journal of Methods in Psychiatric Research* **7**, 33–55.

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

Markon KE, Krueger RF (2006). Information-theoretic latent distribution modeling: distinguishing discrete and continuous latent variable models. *Psychological Methods* **11**, 228–243.

Muthén BO (2004). *Mplus Technical Appendices*. Muthén & Muthén: Los Angeles, CA.

Muthén LK, Muthén BO (2007). *Mplus User's Guide*, 3rd edn. Muthén & Muthén: Los Angeles, CA.

Nylund KL, Asparouhov T, Muthén BO (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: a Monte Carlo simulation study. *Structural Equation Modeling* **14**, 535–569.

O'Connor B (2000). SPSS and SAS programs for determining the number of components using parallel analysis and Velicer's MAP test. *Behavior Research Methods, Instrumentation, and Computers* **32**, 396–402.

Prisciandaro JJ, Roberts JE (2005). A taxometric investigation of unipolar depression in the National Comorbidity Survey. *Journal of Abnormal Psychology* **114**, 718–728.

Preacher KJ, MacCallum RC (2003). Repairing Tom Swift's Electric Factor Analysis Machine. *Understanding Statistics* **2**, 13–43.

Raftery AE (1993). Bayesian model selection in structural equation models. In *Testing Structural Equation Models* (ed. K. A. Bollen and J. S. Long), pp. 163–180. Sage Publications, Inc.: Thousand Oaks, CA.

Robins LN, Wing J, Wittchen H-U, Helzer JE, Babor TF, Burke J, Farmer A, Jablenski A, Pickens R, Regier DA, Sartorius N, Towle LH (1988). The Composite International Diagnostic Interview: an epidemiologic instrument suitable for use in conjunction with different diagnostic systems and in different cultures. *Archives of General Psychiatry* **45**, 1069–1077.

Ruscio AM, Ruscio J (2002). The latent structure of analogue depression: should the Beck Depression Inventory be used to classify groups? *Psychological Assessment* **14**, 135–145.

Ruscio J, Haslam N, Ruscio AM (2006). *Introduction to the Taxometric Method: a Practical Guide*. Erlbaum: Mahwah, New Jersey.

Ruscio J, Ruscio AM (2000). Informing the continuity controversy: a taxometric analysis of depression. *Journal of Abnormal Psychology* **109**, 473–487.

Ruscio J, Zimmerman M, McGlinchey JB, Chelminski I, Young D (2007). Diagnosing major depressive disorder: XI. A taxometric investigation of the categorical–dimensional debate on the structure underlying *DSM-IV* symptoms. *Journal of Nervous and Mental Disease* **195**, 10–19.

SAS Institute, Inc. (2002). %SREGSUB macro provides additional capabilities for PROC SURVEYREG (http://support.sas.com/kb/24/985.html#ref). Accessed 5 June 2008.

Shafer AB (2006). Meta-analysis of the factor structures of four depression questionnaires: Beck, CES-D, Hamilton, and Zung. *Journal of Clinical Psychology* **62**, 123–146.

Slade T, Andrews G (2005). Latent structure of depression in a community sample: a taxometric analysis. *Psychological Medicine* **35**, 489–497.

Solomon A, Ruscio J, Seeley JR, Lewinsohn PM (2006). A taxometric investigation of unipolar depression in a large community sample. *Psychological Medicine* **36**, 973–985.

Spitzer RL, Williams JBW, Gibbon M, First MB (1992). The Structured Clinical Interview for DSM-III-R (SCID). *Archives of General Psychiatry* **49**, 624–636.

Sullivan MD, LaCroix AZ, Russo JE, Walker EA (2001). Depression and self reported physical health in patients with coronary disease: mediating and moderating factors. *Psychosomatic Medicine* **63**, 248–256.

Wacker HR, Battegay R, Muellejans R, Schloesser C (1990). Using the CIDI-C in the general population. In *Psychiatry: A World Perspective*, vol. 1 (ed. C. N. Stefanis, A. D. Rabavilas and C. R. Soldatos), pp. 138–143. Elsevier Science Publishers B. V. (Biomedical Division): New York.

Wade TJ, Cairney J (2000). Major depressive disorder and marital transition among mothers: results from a national panel study. *Journal of Nervous and Mental Disease* **188**, 741–750.

Waller NG, Meehl PE (1998). *Multivariate Taxometric Procedures: Distinguishing Types from Continua*. Sage Publications, Inc.: Thousand Oaks, CA.

Watson D (2003). Investigating the construct validity of the dissociative taxon: stability analyses of normal and pathological dissociation. *Journal of Abnormal Psychology* **112**, 298–305.

Wittchen H-U (1991). Cross-cultural feasibility, reliability and sources of variance of the Composite International Diagnostic Interview (CIDI). *British Journal of Psychiatry* **159**, 645–653.

Wittchen H-U (1994). Reliability and validity studies of the WHO Composite International Diagnostic Interview: a critical review. *Journal of Psychiatric Research* **28**, 57–84.

Zhao S, Kessler RC, Wittchen H-U (1994). Diagnostic algorithms for NCS/DSM-III-R Di. NCS Working Paper no. 7 (http://www.hcp.med.harvard.edu/ncs/ftpdir/working_paper_7.pdf). Accessed 1 September 2006.