Measurement and description of underlying dimensions of comorbid mental disorders using Factor Mixture Models: results of the ESEMeD project

JOSUÉ ALMANSA,¹ JEROEN K. VERMUNT,² CARLOS G. FORERO,¹ GEMMA VILAGUT,^{1,3}
RON DE GRAAF,⁴ GIOVANNI DE GIROLAMO⁵ & JORDI ALONSO^{1,3,6}

1 Health Services Research Unit, IMIM (Hospital del Mar Research Institute), Barcelona, Spain

2 Department of Methodology and Statistics, Faculty of Social and Behavioural Sciences, Tilburg University, Tilburg, The Netherlands

3 CIBER en Epidemiología y Salud Pública (CIBERESP), Barcelona, Spain

4 Trimbos-instituut, Netherlands Institute of Mental Health and Addiction, Utrecht, The Netherlands

5 IRCCS Fatebenefratelli, Brescia, Italy

6 Department of Experimental and Health Sciences, University Pompeu Fabra, Barcelona, Spain

Key words

internalizing disorders,
externalizing disorders,
measurement invariance,
mental comorbidity,
latent classes

Correspondence

Jordi Alonso, Health Services
Research Unit, IMIM (Hospital del
Mar Research Institute), Carrer
del Doctor Aiguader, 88, 08003
Barcelona, Spain.
Telephone (+34) 93 316 07 60
Fax (+34) 93 316 07 97
Email: jalonso@imim.es

Received 28 January 2010;
revised 26 August 2010;
accepted 26 October 2010

Abstract

Epidemiological studies on mental health and mental comorbidity are usually based on prevalences and correlations between disorders, or some other form of bivariate clustering of disorders. In this paper, we propose a Factor Mixture Model (FMM) methodology based on conceptual models aiming to measure and summarize distinctive disorder information in the internalizing and externalizing dimensions. This methodology includes explicit modelling of subpopulations with and without 12 month disorders (“ill” and “healthy”) by means of latent classes, as well as assessment of model invariance and estimation of dimensional scores. We applied this methodology with an internalizing/externalizing two-factor model, to a representative sample gathered in the European Study of the Epidemiology of Mental Disorders (ESEMeD) study – which includes 8796 individuals from six countries, and used the CIDI 3.0 instrument for disorder assessment. Results revealed that southern European countries have significantly higher mental health levels concerning internalizing/externalizing disorders than central countries; males suffered more externalizing disorders than women did, and conversely, internalizing disorders were more frequent in women. Differences in mental-health level between socio-demographic groups were due to different proportions of healthy and ill individuals and, noticeably, to the ameliorating influence of marital status on severity. An advantage of latent model-based scores is that the inclusion of additional mental-health dimensional information – other than diagnostic data – allows for greater precision within a target range of scores.
Copyright © 2011 John Wiley & Sons, Ltd.

Background and aims

Several authors have improved the conceptual framework of mental health with proposals that aim to explain the co-occurrence of mental disorders. Statistically, these models involve explaining the joint presence of psychiatric disorders by modelling their inherent correlations (Krueger, 1999; Krueger and Finger, 2001; Markon and Krueger, 2005; Slade and Watson, 2006; Watson, 2005). In spite of these theoretical developments, univariate and bivariate analysis of descriptive information have been the norm in mental health epidemiology when approaching the study of disorder comorbidity. Although these analysis might be appropriate when the interest resides on specific disorders, such approaches ignore much of the information shared by observed disorders, and thus overlook individual health state. Multivariate models might provide better accounts of mental health description in the population as long as they allow for the joint presence of disorders (i.e. the comorbidity pattern) according to the existent psychiatric knowledge.

Multivariate modelling strategies comprise an additional benefit. A dimensional definition of mental health beyond the classical classification system, such as the International Classification of Diseases (ICD) or the Diagnostic and Statistical Manual of Mental Disorders (DSM) (American Psychiatric Association, 2000; World Health Organization, 1993), is gaining strength among psychiatric researchers. In the study of mental comorbidity it is assumed that individuals' non-observable "psychic state" determines their vulnerability to suffering certain disorders. This vulnerability is also known as diathesis (Clark, 2005). By means of latent variable modelling, it is possible to capture common information shared by the disorders, based on previous psychiatric conceptual research [for example, see Krueger (1999), Mineka *et al.* (1998) and Vollebergh *et al.* (2001)]. Moreover, this information is of dimensional nature. Along these lines, conceptual models of mental health have proposed two broad factors for the expression of this vulnerability: internalizing (depression and anxiety) and externalizing (behavioural and substance disorders) (Cerdeira *et al.*, 2008).

Most desirably, scoring values on the dimensions should allow for fair comparisons between individuals, regardless of their population of origin. Then, group comparisons would be immediate. This property of the scores is known in psychometrics as measurement invariance (MI), and it warrants that the observed variables measure the latent dimension in the same way, regardless of sampled characteristics in each

subpopulation (Mellenbergh, 1989; Meredith, 1993). Although MI is a crucial property for any measurement instrument, it has seldom been tested on the mental-health conceptual models. Non-invariance in a disorder outcome would imply that individuals with the same actual mental-health level would have a differential probability of endorsing the disorder depending on their subpopulation group they belong to. MI only exists if measurement parameters hold equal for any subpopulation. This is clearly the most desirable case, because individuals can be compared based on factor scores only, regardless of their subpopulation of origin. Then, group comparisons based on factor scores are immediate.

Latent variable modelling as a scoring method would allow populations from a conceptual model to be described, because they provide the means to compute individual scores. In the present study we propose the use of latent variable models to answer the requirements of epidemiological studies on mental health models of comorbidity. We emphasize the usefulness of latent modelling not just for conceptual model testing, but also to provide the means to estimate scores of non-observable dimensions of mental health and to contrast whether the measures are invariant. This article also includes a very new approach to the epidemiologic mental-health assessment, consisting on the assumption that there exists a "healthy" subpopulation, that is, that the population includes a proportion of individuals who would not endorse any (12-month) disorder, through a Factor Mixture Model (FMM).

Especially when dealing with observational studies, so common in epidemiology research, the population may display excessive heterogeneity. On certain instances the source of variance might be known, observed and conveniently coded as part of the data gathering process. But in numerous other occasions it is impossible to identify the source of heterogeneity. Extensions of latent variable (factor) models allow individual probabilities pertaining to different, non-observed groups or *latent classes* to be estimated. These models are known as FMMs because they combine continuous latent variables (factors) with discrete latent variables (mixtures or classes). Model parameters in FMMs differ according to the class membership (Lubke and Muthén, 2005, 2007; Lubke and Neale, 2006), and thus, individual scores can be qualified depending on a group of membership (for instance, disorder severity). An important feature is that classes are estimated, and so, instead of using variables to indicate group membership, the mixture models estimate individual probabilities pertaining to each latent class. Differently from stratified or multigroup analysis, where

the grouping variable is explicitly observed, a FMM grouping variable is also latent (i.e. unobserved). This approach has two main advantages. Firstly, they allow unknown sources of variability to be taken into account. Secondly, it is possible to estimate factor scores along with the probability of class membership. Class characteristics can be then explained with covariates such as demographics. It is important to remark that these covariates might be related with the latent class, but the latent class is not defined by them, but by model specifications. Thus, class covariates allow for posterior epidemiologic inferences on class characteristics to identify the source of variability, but the model has already taken those sources into account.

To sum up, FMM allows (a) continuous scores based on categorical information (such as a pattern of disorders or symptoms) to be obtained, (b) confounding sources of heterogeneity to be controlled, (c) covariates to be introduced into the model to explain the latent classes, which is a practical method for epidemiologic inferences, and (d) model-based invariance contrasts.

In the following sections we exemplify how to conduct such analysis using data from the European Study of the Epidemiology of Mental Disorders (ESEMeD), addressing the following goals: (a) to describe the mental health related with internalizing/externalizing disorders of the European population according to their (jointly) observed disorders in the last 12 months, based on the internalizing/externalizing conceptual model; (b) to assess model invariance across socio-demographic groups; (c) to explicitly model the healthy and ill subpopulations jointly.

The internalizing/externalizing disorder conceptual model was chosen because of its simplicity and parsimony, in order to present how this FMM and scoring methodology can be applied, and the new type of information that can be provided. Although more complex models of disorder structure exist, the objective of this paper is not to choose sides on conceptual models. Alternatively, more detailed conceptual models can be also applied like the three factor models (Krueger, 1999; Vollebergh *et al.*, 2001) or more recent models like the six factor structure proposed by Wittchen *et al.* (2009). The modelling strategy would be the same for any other conceptual model, but we deemed the internalizing/externalizing model to be sufficiently simple and yet descriptive to present the FMM methodology. Also, the FMM methodology could be used with conceptual models based on symptom data, not from disorders, or even combining both strategies in a hierarchical model. The procedure would be essentially the same as long as the indicators are considered categorical.

Methods

Sample

The ESEMeD project was a cross-sectional survey based on a stratified, multi-stage, clustered area probability sample. Individuals were assessed in person at their homes using computer-assisted interview (CAPI) techniques. The target population was the non-institutionalized adult population (aged 18 years or older) of Belgium, France, Germany, Italy, the Netherlands and Spain. The sampling frame and the number of sampling stages used to obtain the final sample differed across countries. A more detailed description of the methods and the participants of the ESEMeD project is provided elsewhere (Alonso *et al.*, 2002; Alonso *et al.*, 2004a). In total 21,425 respondents provided data for the project between January 2001 and August 2003. The overall response rate in the six countries studied was 61.2%, with the highest rates in Spain (78.6%) and Italy (71.3%), and lower rates in Germany (57.8%), the Netherlands (56.4%), Belgium (50.6%) and France (45.9%). A two-phase interview procedure was used. The first phase contained the diagnostic assessment of the most common mood and anxiety disorders, health related quality of life, health services utilization and main demographic characteristics. Interviewees who exceeded a predetermined number of symptoms of specific mood or anxiety disorders, and a random 25% of the rest, were asked for the second phase (Part 2 respondents, $N = 8796$). The second phase included an in-depth interview about additional mental disorders (post-traumatic stress disorder and externalizing disorders) and physical chronic conditions amongst other information. Mental disorders assessment was based on version 3.0 of the World Health Organization Composite International Diagnostic Interview (CIDI 3.0) (Kessler and Ustun, 2004), a fully structured, lay administered diagnostic interview that generates diagnoses according to both the ICD-10 and DSM-IV criteria. Within each country, individuals were weighted to adjust for the differential probabilities within household and the differential probabilities of selection for the second phase. Additional post-stratification weights were applied to restore age and gender distribution of the population within each country and the relative dimension of the population across countries. Disorders' descriptive analysis of the ESEMeD data can be found in (Alonso *et al.*, 2004a, 2004b, 2004c). Data for the analyses conducted in this research article consisted on the Part 2 subsample ($N = 8796$).

The main indicator variables analysed in this article are binary, and they describe whether a mental disorder was present in the previous 12 months to the interview (i.e. disorder diagnostic in the last year). Disorder diagnostics

were established by applying DSM-IV criteria to responses to the CIDI 3.0. The diagnoses under analysis included: major depression episode (mde), dysthymia (dys), general anxiety disorder (gad), post-traumatic stress disorder (ptsd), agoraphobia (ago), specific phobia (sp), social phobia (so), panic disorder (pds), alcohol disorders (alc) – including alcohol abuse or alcohol dependence – and behavioural disorders [conduct disorder (cd), oppositional defiant disorder (odd) and attention deficit disorder (add)]. Disorder definitions did not include hierarchy rules. The behavioural disorders are considered as mainly childhood and adolescent disorders, and they were not asked to individuals older than 44 years old because of concerns about recall bias among older respondents (Kessler *et al.*, 2007). Consequently, for the analysis of this article, it was assumed that individuals older than 44 years old had zero-probability to endorse these disorders during the previous 12 months. This assumption is warranted given the decreasing estimated prevalence of having any of these disorders as age increases (for age groups 18–24, 25–34 and 35–44, the 12-month prevalence of having any behavioural disorder were 1.8%, 1.1% and 0.6%, respectively).

Statistical analysis

Confirmatory Factor Analysis (CFA) was used to construct the statistical models from the conceptual internalizing/externalizing disorder models. CFA groups variables in previously defined factor or dimensions, where the observed variables are indicators of the unobserved factors. Because of the categorical nature of our data (binary data: having or not the disorder) a *probit* link was used to relate the continuous latent factor with the binary observed variables, leading to an item response theory (IRT) model: two-parameter normal-ogive model (De Boeck, 2004; Lord and Novick, 1968). The measurement parameters are the intercepts and factor loadings.

First, the internalizing factor was explored, beginning with the simplest model: a single-factor model fitted to the internalizing disorders (mde, dys, ago, pds, gad, pts, so, sp). This is the baseline model.

On a second step, we conducted invariance analyses on the IRT fitted model (Forero and Maydeu-Olivares, 2009), in order to determine if measurement parameter differences among subpopulations could exist. To do so, we used multigroup analyses, which consisted on comparing three different models for each grouping variable: a restricted (invariant) model and two additional unrestricted, non-invariant models. In the invariant model, the factor mean was fixed to zero in one group and freely estimated into the others; thus factor means and variances were free to vary across

groups while measurement parameters held invariant. The first non-invariant unrestricted model allowed the intercepts to differ between groups. The second non-invariant model, allowed both the intercepts and factor loadings to differ between groups. For both unrestricted models, factor means were fixed to zero in each group for identification purposes. Then, for each grouping variable, we assessed whether the unrestricted models significantly differed in model fit as compared to the restricted one (Lubke *et al.*, 2003; Muthén and Asparouhov, 2002). We also compared the restricted model to the baseline model (internalizing single-factor) to assess whether the grouping variable could operate as a covariate for factor scores. Multigroup analysis was conducted according to the following socio-demographic grouping variables: gender, age-group, country, physical chronic condition, years of education, marital status, income, urbanicity and working-status. Due to the low prevalence of the externalizing disorders, which led to estimation problems, the MI analyses were conducted only on the internalizing factor – an MI analyses on the externalizing factor, tested in a larger sample, can be found in Markon and Krueger (2005).

On a third step we developed a mixture CFA/IRT model (FMM), in order to summarize the different solutions obtained in the multigroup approach in a few main classes, instead of having a large number of different solutions (one per group variable), and to explore alternative sources of heterogeneity beyond the observed grouping variables.

As in the multigroup analysis, we tested the FMM versions of the restricted (invariant) and unrestricted (non-invariant) models. Covariates predicting class membership were mainly selected according to those that seemed to have a significant effect in the multigroup analyses. Additionally, we also tested different covariates predicting different levels of severity within the ill class.

Once the factor model was established for internalizing disorders, we included an externalizing factor into the model, constructed from alc, cd, odd and add disorders. The existence of a single externalizing factor is based on conceptual psychiatric information (Krueger, 1999; Krueger *et al.*, 2005; Markon and Krueger, 2005; Slade and Watson, 2006). We also probed for latent classes with different means in the externalizing factor (invariant FMM), and covariates were included to predict class membership probabilities.

Finally, factor scores were estimated for each individual in the final model, which included the appropriate classes and significant covariates. Factor scores were computed as the mean of the factor posterior distribution, along with their standard deviation (Vermunt and Magidson, 2005). This scoring method provides a

measure in a continuous scale for both the internalizing and externalizing health state levels. Disorder variables were coded in the model in a way so that mentally healthy individuals achieved higher factor scores and lower scores represented ill individuals. In this manner, the latent variable stands for an internalizing/externalizing mental health state factor, ranging from worse to better mental health.

Model comparison was mainly made using the Bayesian information criterion (BIC). We also examined other comparative fit indexes for decision-making: Akaike's information criterion (AIC and AIC3 – an AIC modification that heavily penalizes factor over-parameterization). Using these indexes it is possible to choose the best-fitting model among a set of nested or non-nested alternative models (Raftery, 1995). The likelihood ratio test for nested-model comparison was not used in the MI analyses because the large sample size led to significance even with minor differences. Alternatively, model fit was assessed using bivariate residuals (BVRs). The BVR reflects any residual association among pairs of observed after model estimation. Given a model with satisfactory fit, it is expected that no substantial amount of association between every pair of variables is left (Vermunt and Magidson, 2005).

Notice that the election of the internalizing/externalizing two-factor model is just an example of this methodology. More complex models can be implemented and compared using this same approach.

In order to obtain unbiased estimates and standard errors, the complex sample design was taken into account in all analysis by using linearization variance estimator (Alonso *et al.*, 2004a; Vermunt and Magidson, 2005). All models were estimated via maximum likelihood, with 125 start values to avoid a local-optimum solution, using the software LatentGold version 4.5.

Results

Specific phobia and major depression were the most prevalent disorders, with percentages 5.8% and 4.1%, respectively (Table 1). Internalizing disorders were more frequent than externalizing ones. It was estimated that 12.2% of the population have at least one 12-month disorder (Table 2). There existed some tendencies in the prevalence of the disorders. For example, internalizing disorders appeared more frequently in females than in males, but externalizing ones were more frequent in males. Moreover, the youngest individuals had larger prevalences than the oldest ones, and Italy and Spain presented lower disorder prevalences compared to northern countries (see Table 2).

Table 1 Prevalence at 12-month mental disorders weighted

Disorder	N	Prevalence (%)	95% confidence interval
Dysthymia	283	1.23	1.03–1.47
Major depression	905	4.09	3.72–4.49
Agoraphobia	110	0.69	0.50–0.93
General anxiety disorder	196	0.94	0.75–1.17
Panic disorder	174	0.80	0.65–0.98
Post-traumatic stress	214	1.13	0.93–1.38
Social phobia	228	1.62	1.31–2.00
Specific phobia	679	5.77	5.10–6.51
Alcohol disorders	93	0.77	0.52–1.13
Conduct disorder	11	0.07	0.03–0.14
Oppositional defiant disorder	21	0.17	0.09–0.30
Attention deficit disorder	59	0.39	0.24–0.63

Measurement invariance

Table 3 displays goodness-of-fit indices for the internalizing factor models testing MI. The baseline model was compared against the invariant (restricted) ones from the multigroup analyses (baseline model was just the one factor internalizing), and the invariant and non-invariant models against one another. In all cases, the invariant (restricted) model was to be preferred based on the BIC. According to the AIC and AIC3, for some of the grouping variables the free-intercept non-invariant model should be preferred over the invariant one, although their intercepts were not substantially different across groups (parameter estimates provided on request). Given these results, choosing the measurement invariant one is an acceptable election for the sake of parsimony. Regarding the comparisons between the restricted models with the baseline model, it can be seen that the covariates gender, chronic-condition, and country are clearly significant predictors of the internalizing level.

When a latent grouping variable was introduced into the model to construct the FMMs, the appropriate number of latent classes was studied first (Table 4). Models with three classes did not improve significantly their goodness-of-fit respect the two-class models, both in the restricted and unrestricted cases. Fully restricted models (free intercepts and factor loadings) were discarded from subsequent FMMs, given that they did not yield better fit than free-intercept models. The inclusion of class covariates to predict the probability of

Table 2 Prevalence of 12 month type of disorders by socio-demographic variables. Weighted percentages

	Any internalizing		Any externalizing		Any disorder		No disorder		Total	
	N	Percentage	N	Percentage	N	Percentage	N	Percentage	N	Percentage
Gender										
Male	519	7.3	97	1.8	567	8.5	3122	91.5	3689	48.2
Female	1288	15.3	66	0.7	1310	15.6	3797	84.4	5107	51.8
Age										
18–24	173	15.4	32	3.8	185	17.4	479	82.6	664	11.4
25–34	339	10.8	55	2.4	362	12.4	1237	87.6	1599	18.3
35–49	583	12.8	59	1.0	609	13.4	2060	86.6	2669	27.8
20–64	450	11.6	17	0.4	459	11.9	1738	88.1	2197	21.8
+65	262	7.9	0	0.0	262	7.9	1405	92.1	1667	20.7
Country										
Belgium	219	12.6	29	2.0	231	13.2	812	86.8	1043	3.8
France	387	17.7	36	1.9	404	18.9	1032	81.1	1436	20.5
Germany	257	10.0	26	1.5	269	11.0	1054	89.0	1323	31.5
Italy	279	8.7	10	0.3	282	8.8	1497	91.2	1779	22.4
Netherlands	261	12.4	43	2.6	279	13.7	815	86.3	1094	6.1
Spain	404	9.3	19	0.5	412	9.7	1709	90.3	2121	15.6
Chronic condition										
Yes	720	14.0	44	0.7	733	14.3	2184	85.7	2917	30.4
No	1087	10.3	119	1.5	1144	11.3	4735	88.7	5879	69.6
Education										
Up to 12 years	656	11.9	77	1.4	690	12.8	2591	87.2	3281	34.6
More than 12 years	1151	11.2	86	1.1	1187	11.9	4328	88.1	5515	65.4
Work status										
Working	960	11.1	111	1.3	1013	11.9	3850	88.1	4863	56.5
Not working	847	11.8	52	1.2	864	12.6	3069	87.4	3933	43.5
Urbanicity										
Less than 10,000 inhabitants	488	10.8	36	0.9	504	11.3	2021	88.7	2525	33.2
10,000 to 100,000 inhabitants	767	11.5	77	1.1	805	12.3	3035	87.7	3840	38.7
More than 100,000 inhabitants	552	12.0	50	1.8	568	13.1	1863	86.9	2431	28.1
Income										
Low	365	13.9	32	2.5	377	15.6	1213	84.4	1590	19.0
Low average	566	11.1	53	1.0	588	11.5	2121	88.5	2709	32.1
High Average	591	10.5	54	0.9	614	11.2	2362	88.8	2976	33.2
High	285	11.1	24	0.8	298	11.7	1223	88.3	1521	15.7
Marital status										
Married/cohabiting	1109	11.0	84	0.8	1155	11.6	4633	88.4	5788	66.8
Separated/widow/divorced	313	13.1	16	0.5	316	13.2	1011	86.8	1327	11.1
Never married	385	11.9	63	2.9	406	13.6	1275	86.4	1681	22.1
Total	1807	11.4	163	1.2	1877	12.2	6919	87.8	8796	100

class membership widely improved the models fit. Although the two-class model did not improve over the baseline (the internalizing one-factor, without covariates), it can be seen in Table 4 that the two-class model with covariates (models 16 and 17) improved over the one-class with covariates (model 18). BIC values suggested that

the measurement invariant model had better fit in FMMS with covariates. Nevertheless, the AIC and AIC3 suggested that the free-intercepts models had better goodness-of-fit; additionally, intercept parameters were significantly distant between classes (data not shown). Close examination of parameter estimates of the best fitting restricted

Table 3 Goodness-of-fit indices testing measurement invariance (MI) for the internalizing factor

	LogL	BIC	AIC	AIC3	Npar
Baseline internalizing model	-5896.4	11938.0	11824.7	11840.7	16
Gender					
Invariant	-5829.2	11821.9	11694.5	11712.5	18
Free-intercepts	-5817.8	11862.7	11685.7	11710.7	25
Free-intercepts & factor loadings	-5811.7	11914.1	11687.5	11719.5	32
Age					
Invariant	-5872.7	11963.3	11793.3	11817.3	24
Free-intercepts	-5828.7	12129.7	11761.5	11813.5	52
Free-intercepts & factor loadings	-5811.0	12348.6	11782.0	11862.0	80
Country					
Invariant	-5849.6	11935.4	11751.3	11777.3	26
Free-intercepts	-5812.5	12179.0	11747.0	11808.0	61
Free-intercepts & factor loadings	-5794.1	12460.0	11780.1	11876.1	96
Chronic condition					
Invariant	-5880.4	11924.3	11796.8	11814.8	18
Free-intercepts	-5870.9	11968.8	11791.8	11816.8	25
Free-intercepts & factor loadings	-5869.7	12030.0	11803.4	11835.4	32
Education					
Invariant	-5895.1	11953.6	11826.1	11844.1	18
Free-intercepts	-5885.7	11998.5	11821.4	11846.4	25
Free-intercepts & factor loadings	-5882.3	12055.2	11828.6	11860.6	32
Income					
Invariant	-5885.6	11971.0	11815.2	11837.2	22
Free-intercepts	-5873.1	12136.7	11832.2	11875.2	43
Free-intercepts & factor loadings	-5864.1	12309.5	11856.2	11920.2	64
Marital status					
Invariant	-5888.4	11958.4	11816.8	11836.8	20
Free-intercepts	-5863.4	12035.5	11794.7	11828.7	34
Free-intercepts & factor loadings	-5850.4	12136.8	11796.9	11844.9	48
Urbanicity					
Invariant	-5894.3	11970.2	11828.5	11848.5	20
Free-intercepts	-5880.3	12069.3	11828.5	11862.5	34
Free-intercepts & factor loadings	-5873.8	12183.6	11843.7	11891.7	48
Work status					
Invariant	-5891.6	11946.7	11819.2	11837.2	18
Free-intercepts	-5884.0	11995.1	11818.0	11843.0	25
Free-intercepts & factor loadings	-5879.3	12049.2	11822.6	11854.6	32

and unrestricted models (models with four covariates: gender, age, country and chronic-condition), gave further evidence of class constitution. Firstly, the second class of the restricted model had a mean around 10.8 and close to zero variance. Secondly, class two of the unrestricted (free-intercepts) model had much larger intercepts than the first class and its variance was close to zero. Thus, in both models, individuals in class two had low probabilities of endorsing any disorder. These results suggest that both the unrestricted (free intercept) and restricted two-classes models revealed “healthy” and “ill” classes – which was

one of the main goals of this paper. The unrestricted model was not better in terms of fit and the class interpretation was the same in both cases, thus the restricted model was chosen for subsequent models.

Since a variance of zero is a boundary solution, the variance of Class 2 was set to zero; furthermore, the class mean was set to 10. Class 2 represents a homogeneous group of healthy individuals, where *homogeneous* denotes no variability in their health status, and *healthy* denotes having a high factor score. This class is clearly different from the class of ill subjects (whose members are *heterogeneous*,

Table 4 Goodness-of-fit indices testing FMMs for the internalizing factor, and final model for internalizing and externalizing factors

		LogL	BIC	AIC	AIC3	Npar	
<i>Internalizing model</i>							
1	Baseline internalizing model	−5896.4	11938.0	11824.7	11840.7	16	
2-Classes							
2	Invariant	−5893.8	11960.2	11825.6	11844.6	19	
3	Free-intercepts	−5857.9	11952.0	11767.8	11793.8	26	
4	Free-intercepts & loadings	−5854.1	12008.0	11774.3	11807.3	33	
3-Classes							
5	Invariant	−5892.0	11983.8	11828.0	11850.0	22	
6	Free-intercepts	−5851.2	12029.3	11774.4	11810.4	36	
7	Free-intercepts & loadings	−5839.3	12132.7	11778.6	11828.6	50	
2-Classes <i>Class membership covariates</i>							
8	Invariant	gender	−5824.5	11830.7	11689.0	11709.0	20
9	Free-intercepts		−5803.4	11852.0	11660.8	11687.8	27
10	Invariant	gender + age	−5792.6	11803.1	11633.1	11657.1	24
11	Free-intercepts		−5768.1	11817.8	11598.3	11629.3	31
12	Invariant	gender + age + country	−5758.0	11779.5	11574.1	11603.1	29
13	Free-intercepts		−5728.8	11784.5	11529.5	11565.5	36
14	Invariant	gender + age + country + chronic	−5729.3	11731.1	11518.7	11548.7	30
15	Free-intercepts		−5705.2	11746.4	11484.3	11521.3	37
16	Invariant (healthy-class restrictions ^a)	gender + age + country + chronic	−5729.3	11713.0	11514.7	11542.7	28
17		+ marital status (<i>severity covariate</i>)	−5720.5	11713.5	11501.0	11531.0	30
18	1-Class	gender + age + country + chronic	−5740.9	11726.9	11535.7	11562.7	27
<i>Internalizing and externalizing model^b Externalizing class-covariates</i>							
19		—	−6419.9	13203.1	12919.8	12959.8	40
20		gender	−6400.5	13173.4	12883.0	12924.0	41
21		gender + age ^c	−6358.6	13107.6	12803.1	12846.1	43
22 (Final model)		gender + age ^c + country	−6343.1	13122.2	12782.3	12830.3	48
23		gender + age ^c + country + chronic	−6343.1	13131.2	12784.2	12833.2	49

^aSecond class restrictions: mean = 10 and variance = 0.

^bInternalizing factor modelled as in model 17. Two-class externalizing factor with (healthy) second class restrictions: mean = 10 and variance = 0.

^cCoded in three categories (18–34, 35–49 and +50). Age covariate in the internalizing factor had five categories (18–24, 25–34, 35–49, 50–64 and +65).

i.e. they do exhibit variability). It follows that the final model combining these two classes is a CFA-type model for Class 1 (mentally ill subjects) whereas all members of Class 2 have been assigned the same high-value in the factor, which implies zero-probability of endorsing every disorder.

Covariates predicting different levels of severity within ill-class individuals were also tested and marital status covariate resulted significant.

Final internalizing/externalizing model

First we fitted a single externalizing factor and we found that it properly explained the four externalizing disorders

in the ESEMeD data because all BVRs among them were very low. Then, the externalizing factor was added to the later internalizing model previously described (model 17 in Table 4). Similarly, a new categorical latent variable was included to describe externalizing healthy and non-healthy population simultaneously (with zero variance and mean 10 restrictions for the healthy class). Both class variables, for the internalizing and externalizing factors, led to a four-class cross-classification. Covariance between internalizing and externalizing factors was estimated only for those individuals belonging to both ill classes, otherwise was fixed to zero. Variables gender, age and country were found to be significant covariates for the externalizing classes (see

Figure 1 Final model-structure for the analysis of internalizing and externalizing mental-health states in the European general population.

Table 5 Measurement parameters

		Factor loading	Intercept
Internalizing factor	Dysthymia	1	2.48 (0.21)
	Major depression	1.41 (0.21)	1.94 (0.26)
	Agoraphobia	0.84 (0.22)	2.62 (0.33)
	General anxiety disorder	0.95 (0.20)	2.58 (0.27)
	Panic disorder	0.75 (0.18)	2.44 (0.20)
	Post-traumatic stress	0.51 (0.11)	2.03 (0.15)
	Social phobia	0.39 (0.12)	1.77 (0.14)
	Specific phobia	0.21 (0.08)	0.98 (0.10)
Externalizing factor	Alcohol disorders	1	1.85 (0.26)
	Conduct disorder	1.09 (0.40)	3.01 (0.22)
	Oppositional defiant	2.90 (2.46)	4.21 (2.35)
	Attention deficit	2.03 (1.39)	2.81 (1.12)

Note: Standard errors given in parentheses.

models 19 to 23 in Table 4). BIC results for covariate country indicated no improvement in goodness-of-fit, but AIC and AIC3 supported the hypothesis of better fit. In addition, country parameters were significantly different from zero. The covariate age on the externalizing factor was recoded into three categories (18–34, 35–49 and more than 50 years old), instead of the five categories used in most of the ESEMeD data analyses, because of the very low prevalence of externalizing disorders in older population.

The structure of the final model (model 22 in Table 4) is shown in Figure 1. Tables 5–7 show the parameter estimation of this model.

Parameter covariates indicated that older people and southern countries (Spain and Italy) had lower probability of belonging to the ill class; males were more likely than females to belong to the externalizing ill class, but less likely to belong to the internalizing ill class. Having any physical chronic condition increase the probability of belonging to the internalizing ill class (Tables 5–7). The effect of gender, age, country and chronic-condition on the internalizing/externalizing dimensions was mediated by the latent classes, and they did not explained differences in the mental-health level within classes. Marital status had a significant effect on the severity of

Table 6 Covariate parameters

Internalizing			Externalizing		
Intercept		0.47 (0.29)	Intercept		2.14 (0.50)
Gender:	Male	0.60 (0.07)	Gender:	Male	−0.92 (0.17)
	Female	−0.60 (0.07)		Female	0.92 (0.17)
<i>Class membership^a</i>					
Age group:	18–24	−0.74 (0.26)	Age group:	18–34	−1.68 (0.27)
	25–34	−0.12 (0.13)		35–49	−0.04 (0.24)
	35–49	−0.32 (0.12)		50+	1.71 (0.33)
	50–64	0.07 (0.14)			
	65+	1.11 (0.18)			
Country:	Belgium	−0.24 (0.20)	Country:	Belgium	−0.90 (0.44)
	France	−0.72 (0.13)		France	−0.32 (0.34)
	Germany	0.20 (0.14)		Germany	−0.32 (0.33)
	Italy	0.42 (0.11)		Italy	1.64 (0.45)
	Netherland	−0.03 (0.14)		Netherland	−1.35 (0.38)
	Spain	0.35 (0.13)		Spain	1.25 (0.44)
Chronic condition:	Yes	−0.49 (0.09)			
	No	0.49 (0.09)			
<i>Severity within ill class^b</i>					
Marital Status					
Married/Cohabiting		0.18 (0.06)			
Previously married		−0.19 (0.09)			
Never married		0.01 (0.08)			

Note: Standard errors given in parentheses.

^aLogistic regression parameters: (dependent) class variable used Class 1 (ill) as reference; (independent) covariates used effect coding.

^bLinear regression parameters: internalizing mean differences across covariate categories, within Class 1.

Table 7 Factor mean and variances/covariances

Class ^a		Internalizing-factor		Externalizing-factor		Factor-covariance
Internalizing	Externalizing	Mean	Variance	Mean	Variance	
1	1	0	1.02 (0.25)	0	0.47 (0.27)	0.60 (0.28)
1	2	0	1.02 (0.25)	10	0	0
2	1	10	0	0	0.11 (0.13)	0
2	2	10	0	10	0	0

Note: Standard errors given in parentheses.

^aClass 1 refers the ill-class. Class 2 refers to the healthy class.

the internalizing mental health among those who had internalizing disorders: married (or cohabiting) individuals showed better internalizing health, but previously married (separated, divorced or widowed) had a lower internalizing level.

Both the externalizing and internalizing factor yielded two classes (healthy and ill); consequently, subjects may fall

into any of four possible class-combinations. The estimated covariance factor in subjects belonging simultaneously to the classes “externalizing-ill” and “internalizing-ill” (i.e. comorbid individuals) was 0.60, leading to a factor-correlation of 0.87, whereas the externalizing variance estimated for the combination of “externalizing-ill” and “internalizing-healthy” was very low (see Table 7). This

result has two important implications: (a) among mentally comorbid subjects, both dimensions are highly correlated and (b) all subjects suffering of externalizing disorders only (i.e. “externalizing-ill” and “internalizing-healthy”) had very similar levels of mental health regarding the externalizing dimension.

About the goodness-of-fit of this final model, most of the BVRs were low, so the model can reproduce the observed association between pairs of the disorder variables quite well. Only few of them indicated somewhat significant residual associations between certain pairs of anxiety disorders: specific phobia with agoraphobia (19.9), specific with social phobia (10.2) and social phobia with agoraphobia (12.8).

Factor scores

Factor scores on the final model were estimated as posterior factor means, therefore taking also into account the posterior class membership probability (Vermunt and Magidson, 2005). Table 8 shows the maximum and minimum probabilities of belonging to Class 1 and factor scores, by disorder type. Recall that ill individuals formed Class 1 (zero factor mean) whereas Class 2 is formed by healthy individuals (factor mean 10 and zero variance). All individuals diagnosed with any internalizing disorder were assigned for sure (with probability one) to the internalizing ill class – except the individuals with only specific phobia, who had a probability to belong to the ill class ranging from 0.9032 to 0.9998, and some of them had a factor score slightly over zero. Individuals with at least one externalizing disorder were assigned with probability one to the externalizing ill class, all of them with negative factor scores. Only two individuals without any internalizing disorder (but with odd and add externalizing disorders) scored -0.38 and 0.52 in the internalizing factor, lower than the maximum score

for individuals with any internalizing disorder (1.16). A part from these specific two cases, individuals without disorders scored positively in both dimensions and their internalizing (or externalizing) scores were higher than those individuals with any internalizing (or externalizing) disorder.

There were cases of individuals without any disorder but with a high probability of belonging to Class 1, up to 0.94 and 0.85 for internalizing and externalizing factor respectively (Table 8). Figure 2 shows the histogram of the score distribution by classes, once the individuals have been assigned to their highest probability class, for the internalizing (Figure 2a) and externalizing factor (Figure 2b).

It is possible to conclude from the score distributions that externalizing disorders were much less prevalent than the internalizing ones. The estimated latent class sizes were: for the internalizing factor the Class 1 included 34.7% of individuals and Class 2 a 65.3%; and for the externalizing factor the Class 1 included 16.2% of the individuals and Class 2 a 83.8%. According to the model, it is expected 21.6% of misclassification in the internalizing factor and 11.6% in the externalizing factor when assigning the individuals to their most likely class.

Figure 3 displays the bivariate plot of the factor scores. Dotted lines mark the maximum observed scores for those with any disorder. Average scores for age-group, country and gender categories have been marked in Figure 3 (see also Table 9). It can be seen in Figure 3 the high correlation on the group of individuals having externalizing and internalizing disorders, and the low variability in the group with only externalizing disorders.

Figure 3 also makes evident a tendency for younger people to suffer from worse internalizing/externalizing mental health state, disorder differences by gender (males are more prevalent in externalizing disorders and females in the internalizing ones), and southern countries showed

Table 8 Minimum and maximum values of probability class membership and factor scores

	Internalizing		Externalizing	
	Probability ill-class ^a	Score	Probability ill-class ^a	Score
Any internalizing disorder				
Yes	[0.90, 1.00]	[−2.96, 1.16]	[0.00, 1.00]	[−1.67, 10.0]
No	[0.03, 0.94]	[−0.38, 9.74]	[0.00, 1.00]	[−1.26, 9.98]
Any externalizing disorder				
Yes	[0.09, 1.00]	[−2.52, 9.09]	[1.00, 1.00]	[−1.67, −0.23]
No	[0.03, 1.00]	[−2.96, 9.74]	[0.00, 0.85]	[1.43, 10.0]

^aMembership probability to belong to the ill class (Class 1).

Figure 2 Histogram of the estimated scores from the internalizing (A) and externalizing (B) mental-health dimensions, stratified by class membership (individuals assigned to their highest probability class).

better levels of internalizing/externalizing mental health. Socio-demographic mean scores were quite high because most of the population was classified as mentally healthy. Table 9 provides a more detailed summary of class membership and factor scores. Notice that factor means for the comorbid class were significantly lower than factor means in groups with just one disorder: internalizing and externalizing averages for the comorbid group were -1.28 and -0.95 , respectively, while the internalizing averages in the pure internalizing group is -0.48 and -0.41 in the pure externalizing. The lower average scores estimated for any-disorder respect pure-disorder can be explained by this negative interaction in the comorbid group. Differences in the externalizing outcomes for chronic condition and marital status groups are explained by their relation with age (having any chronic condition and being previously married is associated with older ages).

Score estimates for diagnosed individuals (i.e. low factor scores) were considerably accurate, as shown by their small standard deviations. Similarly, the very high scores (corresponding to healthy individuals) were estimated fairly precisely. However, middle-score individuals had very large variability, so the model is expected to be imprecise when estimating middle-range scores for both factors (Figure 4).

Discussion

For quite a long time now, there has been a claim in psychiatry research for dimensional continuous measures of mental health states (Brown *et al.*, 1998; Clark and Watson, 1991; Clark, 2005; Krueger, 1999; Krueger and Finger, 2001; Krueger *et al.*, 2005; Mineka *et al.*, 1998; Vollebergh *et al.*, 2001; Watson *et al.*, 1988; Watson, 2005). In spite of the existence of dimensional conceptual models, they had been scarcely used to score and classify the individuals given their comorbidity pattern. In this study, we developed a statistical model for the analysis of population mental health regarding internalizing/externalizing disorders that yielded such measures. This model summarized 12 dichotomous disorders into two continuous variables measuring severity, while the latent class part of the model retained the ability to describe categorical properties of certain subpopulations. Analyses support the usefulness of latent variable models to estimate scores of non-observable internalizing/externalizing mental health states directly from the conceptual models. Thus, dimensional measures with categorical description embedded in the final scores could be attained with relative ease.

To our knowledge, this is the first time that mentally healthy and ill subpopulation have been jointly analysed and explicitly modelled. From a methodological point of view, joint modelling is compulsory when analysing general population, not only because of the improvements in goodness-of-fit, but also because in so doing, the models approach closer to reality. Our two-class model arose without imposing previous constraints, so it turns out to be a most trustworthy approximation to model the data. One advantage of the FMM is that it allows the simultaneous classification of variables (in factors) and individuals (in classes).

While univariate and bivariate analysis of mental disorders are appropriate for detailed variable descriptions, multivariate finite mixture models based on conceptual psychiatric knowledge make available comprehensive depictions of mental health states. In addition, and contrary to the dichotomous diagnostics of disorders,

Figure 3 Bivariate scores grouped by latent classes: individuals assigned to their highest probability 4-class-combination; labels allocate mean scores for covariates age (18–24, 25–34, 35–49, 50–64, 65+), countries and gender, and for those with pure internalizing disorder (INTERN), pure externalizing disorder (EXTERN), both types of disorders (INT&EXT) and no disorder (HEALTHY); dotted lines mark the maximum observed scores for those with any disorder, for each dimension.

latent variables provide continuous and dimensional measures of mental health state. Moreover, factor scores capture the simultaneous effects of a number of observed disorders on the individuals. In this fashion, these models are able to seize interactions and the differential impact of disorders on mental factors. Given the correlation between internalizing and externalizing factors, when no internalizing disorders were observed a relatively low internalizing score was estimated if externalizing disorders were present (and the other way round). This methodology models the comorbidity directly.

Another advantage of FMMs is that they make available a direct method for interpreting results (i.e. scores) across sample characteristics. A graphical representation of dimensional scores and the allocation of average values of covariates can be used as simple procedure to detect the existing mental health (most relevant) tendencies. For instance, in our model it was straightforward to detect differences in the type of disorders depending on gender, age and country of origin. Perhaps more interestingly, it was easy to see that the presence of internalizing/externalizing comorbidity had greater impact on the overall mental health state than pure disorders' type, and

it can be measured quantitatively and directly in terms of level of mental-health state.

Other important benefit of this statistical model is that it measures the effect of the covariates over the common disorders' information, more parsimonious and easily interpretable than having an effect over each one of the disorders. The model also gives an interpretation about how socio-demographic variables influence the internalizing/externalizing dimensions. In this study, we found that the differential mental-health status between gender, age, country and chronic-condition socio-demographic groups was just due to differences in the proportion of healthy and ill individuals, but differences in the illness severity across groups was not found. Conversely, marital status did not show an influence in the proportion of healthy and ill individuals, but it had a significant effect on the severity of internalizing "ill" individuals, in which previously married had lower internalizing level and married (or cohabiting) individuals had better internalizing mental health. This double covariate-effect interpretation (on population proportions and severity) is only possible within the FMM framework.

Table 9 Mean of ill-class membership probability and factor scores

	Internalizing		Externalizing	
	Probability ill-class ^a	Mean	Probability ill-class ^a	Mean
Overall sample	0.35	6.56	0.16	8.38
Any disorder	0.95	−0.08	0.21	7.79
Pure internalizing disorder	0.99	−0.48	0.12	8.73
Any internalizing disorder	0.99	−0.52	0.16	8.33
Pure externalizing disorder	0.32	6.59	1.00	−0.41
Any externalizing disorder	0.59	3.54	1.00	−0.62
Internalizing and externalizing comorbid	1.00	−1.28	1.00	−0.95
No disorder	0.26	7.49	0.15	8.46
Gender				
Male	0.23	7.77	0.26	7.43
Female	0.46	5.44	0.07	9.26
Age group				
18–24	0.46	5.44	0.37	6.30
25–34	0.33	6.71	0.37	6.35
35–49	0.39	6.18	0.14	8.56
50–64	0.36	6.41	0.03	9.66
+65	0.23	7.72	0.03	9.75
Country				
Belgium	0.39	6.10	0.28	7.24
France	0.51	4.92	0.21	7.90
Germany	0.32	6.89	0.20	8.04
Italy	0.27	7.34	0.05	9.48
Netherlands	0.37	6.35	0.36	6.36
Spain	0.29	7.14	0.08	9.18
Chronic condition				
Yes	0.42	5.80	0.08	9.24
No	0.31	6.90	0.20	8.00
Marital status				
Married/cohabiting	0.35	6.60	0.13	8.69
Previously married	0.34	6.56	0.06	9.38
Never married	0.35	6.46	0.31	6.93

^aMembership probability to belong to the ill class (Class 1).

Moreover, greater score precision could be achieved using additional disorder information, other than diagnostic data, such as ordinal information of disorder severity, to increase the reliability of the scores within mild to moderate range of mental health states. Notice that one important advantage of latent model-based scores resides in that the score precision is a function of the score itself and not an instrument's property in a certain sample (Hambleton and Rogers, 1991; De Boeck, 2004). While diagnostic data can only provide information about the most severe mental health states (i.e. those that have been already diagnosed), latent methods allow to add further precision in certain score levels just by taking into

account any other dimensional information (for instance, symptoms or questionnaire data) at any measurement level into the model.

The dimensional modelling in a measurement invariant FMM is done exclusively through continuous latent factors. In our methodology, latent classes only serve to determine the factor mean scores. In this way, latent classes are not sensitive to inclusion or exclusions of disorders, as they are being used to model sample characteristics and not latent score structures.

This model also allows for the assessment of MI. MI is a crucial property for any measurement instrument that researchers overlook very often. Model stability of the

Figure 4 Precision of the internalizing (A) and externalizing (B) estimated scores.

model has seldom been tested in the literature of conceptual mental health models. To our knowledge only Chorpita (2002) conducted measurement invariance analyses in a tripartite model based on symptom's information and Markon and Krueger (2005) on the externalizing factor. There is motive of concern: when MI is not present, the model cannot ensure its applicability over heterogeneous samples. Only a measurement-invariant model is capable to summarize simultaneously all variable information into factor scores. As a result, applied researchers can compare sample characteristics via average scores. The MI property even warrants that the interpretation of the observed data (for example the Table 2 results) is not influenced by unknown confounder effects.

Model limitations

A first limitation is due to the participation rate, which was over 61%. This might limit external validity of our

data. Nevertheless, internal validity is assured by the instruments, measurement and control methods used. And therefore we are totally confident on the validity and relevance of the associations described in this paper.

Another limitation concerning data gathering is that information based on disorders could be acceptable in clinical samples, but it is not good enough over general population when the main purpose is to score mental health states in any kind of disorders. In our example, we found that internalizing/externalizing scores' estimation had unacceptable high variance on the middle-level health states. Krueger and Finger (2001) and Markon and Krueger (2005) found similar results using disorder information. In their studies, factor score precision was only acceptable for ill-range scores, but estimates were more unstable within the healthy range. In our study, the inclusion of latent classes for the healthy and non-healthy subpopulations improved precision of the highest scores, but middle-ranked scores were still inaccurate. Our model was able to discriminate the clear cases of ill and healthy individuals but it was less precise on the intermediate scores, which could represent subthreshold or mild phases of the disorders. This lack of precision produces a 21.6% classification error when assigning individuals to their estimated higher probability internalizing class. Simply put, the model had not enough power to distinguish states in the mild to healthy range. However, those who fulfilled the criteria for at least one disorder were assigned to the ill class, together with those with a mild health state or subthreshold (not so severe to fulfil some disorder criteria). Consequently, the model ill-class refers to a wide range of mental health states. This is a direct consequence of the nature of disorder-level data, in which having the disorder implies the fulfilment of several (strict) criteria, but not having the disorder applies to a wide range of possibilities: from individuals who meet all the criteria except one, to very healthy individuals.

Overall, the FMM performed well in both the internalizing and the externalizing dimension. However, in both dimensions there were a group of individuals with middle score values that were most likely to belong to Class 1. This points to the feasibility of a model with three classes: diagnosed (scores around zero or less), mild (scores approximately from two to five) and healthy (scores over five). On this instance, ill classes would include actual disorders as well as subthreshold or mild mental disorders. We opted for a two-class model because the inclusion of a third class did not substantially improve the BIC and AIC indexes. So, according to the results, a three-class model might explain better the population

mental health state, but such a model is not easy to estimate with only disorder information. Yet, the possibility of expanding the number of classes remains open for future research. Future investigations should search for additional variables measuring mild (and positive) health states to model a sensitive three-class model. This strategy will be undoubtedly fruitful to increase the precision of score estimates all along the scoring continuum.

The internalizing factor captures most of the information provided by the internalizing disorders. Nevertheless, a few significant BVRs between some anxiety disorders remained. The pattern disorder consisting on having specific phobia and not having any other disorder did not result in 100% probability of belonging to internalizing Class 1, as expected from every individual with at least one internalizing disorder. Nonetheless, the estimated probability was quite close to one (at least 0.9). This is a caveat for model interpretation, as the effect of some anxiety disorders on the internalizing score could be slightly underestimated. There are previous results in the literature indicating that anxiety disorders are not totally well explained by a single internalizing factor (Fergusson *et al.*, 2006). Other models can be proposed to overcome this limitation. A two-factor model, splitting internalizing disorders into distress (mde, dys, gad, pts) and fear (pds, sp, so, ago) captured better all the disorder associations (Krueger, 1999; Slade and Watson, 2006; Vollebergh *et al.*, 2001). Still, a single factor construct is useful as a measure of mental health in large population epidemiological research, because it yields small measurement error and does not invalidate its use for epidemiological purposes given its parsimony.

Conclusion

This paper provides a methodology to assess the internalizing and externalizing psychiatric dimensions. This methodology is suitable for epidemiologic studies, and offers easily interpretable results. Once established a nosology model, a set of factors obtained, by means of a latent variable model, summarize the population health state. The model gathers the observed binary information to provide a continuous score estimate for each mental health factor. Whenever the purpose is the population mental-health state assessment, this methodology offers more and clearer information than univariate and bivariate disorders' descriptives. Latent variable models can account for mental comorbidity: the estimated factor scores take into consideration the number of disorders

and their weight in mental-health state. A major contribution of our model is that it permits to describe healthy and ill subpopulations, estimated simultaneously with factor measurement parameters.

Beyond plausibility assumption that a completely healthy psychiatric subpopulation exists, the inclusion of a latent class variable for different levels of severity improved the epidemiological description and increased the score precision on the mental health factor continuum. As recent proposals extend the factor structure of disorders in various ways, FMM provides a method to open a promising line on these models which would allow establishing individual scores, contrasting model adequacy and results from both clinical and epidemiologic approaches.

Acknowledgements

The ESEMeD project (<http://www.epremed.org>) was funded by the European Commission (Contracts QLG5-1999-01042; SANCO2004123), the Piedmont Region (Italy), Fondo de Investigación Sanitaria, Instituto de Salud Carlos III, Spain (FIS 00/0028-02), Ministerio de Ciencia y Tecnología, Spain (SAF 2000-158-CE), Departament de Salut, Generalitat de Catalunya, Spain, and other local agencies and by an unrestricted educational grant from GlaxoSmithKline. ESEMeD is carried out in conjunction with the World Health Organization World Mental Health (WMH) Survey Initiative. We thank the WMH staff for assistance with instrumentation, fieldwork, and data analysis. These activities were supported by the US National Institute of Mental Health (R01MH070884), the John D. and Catherine T. MacArthur Foundation, the Pfizer Foundation, the US Public Health Service (R13-MH066849, R01-MH069864, and R01 DA016558), the Fogarty International Centre (FIRCA R03-TW006481), the Pan American Health Organization, the Eli Lilly & Company Foundation, Ortho-McNeil Pharmaceutical, Inc., GlaxoSmithKline, Bristol-Myers Squibb, and Shire. A complete list of WMH publications can be found at <http://www.hcp.med.harvard.edu/wmh/>.

This specific work was also supported by a research grant from "Comissionat per a Universitats i Recerca del Departament d'Innovació, Universitats i Empresa de la Generalitat de Catalunya i del Fons Social Europeu" (AGAUR: 2009FIC 00015), a grant to support research from Fondo de Investigación Sanitaria, ISCIII (ECA07/059), and a grant from Ministerio de Ciencia e Innovación FSE (JCI-2009-05486).

Declaration of interest statement

All authors have no competing interests.

References

- Alonso J., Angermeyer M.C., Bernert S., Bruffaerts R., Brugha T.S., Bryson H., Girolamo G., Graaf R., Demyttenaere K., Gasquet I., Haro J.M., Katz S.J., Kessler R.C., Kovess V., Lepine J.P., Ormel J., Polidori G., Russo L.J., Vilagut G., Almansa J., Arbabzadeh-Bouchez S., Autonell J., Bernal M., Buist-Bouwman M.A., Codony M., Domingo-Salvany A., Ferrer M., Joo S.S., Martinez-Alonso M., Matschinger H., Mazzi F., Morgan Z., Morosini P., Palacin C., Romera B., Taub N., Vollebergh W.A. (2004a) Sampling and methods of the European Study of the Epidemiology of Mental Disorders (ESEMeD) project. *Acta Psychiatrica Scandinavica*, **109**(420), 8–20.
- Alonso J., Angermeyer M.C., Bernert S., Bruffaerts R., Brugha T.S., Bryson H., de Girolamo G., Graaf R., Demyttenaere K., Gasquet I., Haro J.M., Katz S.J., Kessler R.C., Kovess V., Lepine J.P., Ormel J., Polidori G., Russo L.J., Vilagut G., Almansa J., Arbabzadeh-Bouchez S., Autonell J., Bernal M., Buist-Bouwman M.A., Codony M., Domingo-Salvany A., Ferrer M., Joo S.S., Martinez-Alonso M., Matschinger H., Mazzi F., Morgan Z., Morosini P., Palacin C., Romera B., Taub N., Vollebergh W.A. (2004b) Prevalence of mental disorders in Europe: results from the European Study of the Epidemiology of Mental Disorders (ESEMeD) project. *Acta Psychiatrica Scandinavica*, **109**(420), 21–27.
- Alonso J., Angermeyer M.C., Bernert S., Bruffaerts R., Brugha T.S., Bryson H., Girolamo G., Graaf R., Demyttenaere K., Gasquet I., Haro J.M., Katz S.J., Kessler R.C., Kovess V., Lepine J.P., Ormel J., Polidori G., Russo L.J., Vilagut G., Almansa J., Arbabzadeh-Bouchez S., Autonell J., Bernal M., Buist-Bouwman M.A., Codony M., Domingo-Salvany A., Ferrer M., Joo S.S., Martinez-Alonso M., Matschinger H., Mazzi F., Morgan Z., Morosini P., Palacin C., Romera B., Taub N., Vollebergh W.A. (2004c) 12-Month comorbidity patterns and associated factors in Europe: results from the European Study of the Epidemiology of Mental Disorders (ESEMeD) project. *Acta Psychiatrica Scandinavica*, **109**(420), 28–37.
- Alonso J., Ferrer M., Romera B., Vilagut G., Angermeyer M., Bernert S., Brugha T.S., Taub N., McColgen Z., De Girolamo G., Polidori G., Mazzi F., De Graaf R., Vollebergh W.A., Buist-Bouwman M.A., Demyttenaere K., Gasquet I., Haro J.M., Palacin C., Autonell J., Katz S.J., Kessler R.C., Kovess V., Lepine J.P., Arbabzadeh-Bouchez S., Ormel J., Bruffaerts R. (2002) The European Study of the Epidemiology of Mental Disorders (ESEMeD/MHEDEA 2000) project: rationale and methods. *International Journal of Methods in Psychiatric Research*, **11**(2), 55–67.
- American Psychiatric Association. (2000) *Diagnostic and Statistical Manual of Mental Disorders (DSM-IV)* (4th ed.), Arlington, VA, American Psychiatric Association.
- Brown T.A., Chorpita B.F., Barlow D.H. (1998) Structural relationships among dimensions of the DSM-IV anxiety and mood disorders and dimensions of negative affect, positive affect, and autonomic arousal. *Journal of Abnormal Psychology*, **107**(2), 179–192.
- Cerda M., Sagdeo A., Galea S. (2008) Comorbid forms of psychopathology: key patterns and future research directions. *Epidemiologic Reviews*, **30**, 155–177.
- Chorpita B.F. (2002) The tripartite model and dimensions of anxiety and depression: an examination of structure in a large school sample. *Journal of Abnormal Child Psychology*, **30**(2), 177–190.
- Clark L.A. (2005) Temperament as a unifying basis for personality and psychopathology. *Journal of Abnormal Psychology*, **114**(4), 505–521.
- Clark L.A., Watson D. (1991) Tripartite model of anxiety and depression: psychometric evidence and taxonomic implications. *Journal of Abnormal Psychology*, **100**(3), 316–336.
- De Boeck P. (2004) *Explanatory Item Response Models: A Generalized Linear and Nonlinear Approach*, Berlin, Springer-Verlag.
- Fergusson D.M., Horwood L.J., Boden J.M. (2006) Structure of internalising symptoms in early adulthood. *The British Journal of Psychiatry*, **189**(6), 540–546.
- Forero C.G., Maydeu-Olivares A. (2009) Estimation of IRT graded response models: limited versus full information methods. *Psychological Methods*, **14**(3), 275–299.
- Hambleton R.K., Rogers L. (1991) *Fundamentals of Item Response Theory*, London, Sage Publications.
- Kessler R.C., Üstün T.B. (2004) The World Mental Health (WMH) Survey Initiative Version of the World Health Organization (WHO) Composite International Diagnostic Interview (CIDI). *International Journal of Methods in Psychiatric Research*, **13**(2), 93–121.
- Kessler R.C., Angermeyer M., Anthony J. C., de Graaf R., Demyttenaere K., Gasquet I., de Girolamo G., Gluzman S., Gureje O., Haro J.M., Kawakami N., Karam A., Levinson D., Medina-Mora M.E., Oakley Browne M.A., Posada-Villa J., Stein D.J., Adley Tsang C.H., Aguilar-Gaxiola S., Alonso J., Lee S., Heeringa S., Pennell B.E., Berglund P., Gruber M.J., Petukhova M., Chatterji S., Üstün T.B. (2007) Lifetime prevalence and age-of-onset distributions of mental disorders in the World Health Organization's World Mental Health Surveys. *World Psychiatry*, **6**(3), 168–176.
- Krueger R.F. (1999) The structure of common mental disorders. *Archives of General Psychiatry*, **56**(10), 921–926.
- Krueger R.F., Finger M.S. (2001) Using item response theory to understand comorbidity among anxiety and unipolar mood disorders. *Psychological Assessment*, **13**(1), 140–151.
- Krueger R.F., Markon K.E., Patrick C.J., Iacono W.G. (2005) Externalizing psychopathology in adulthood: a dimensional-spectrum conceptualization and its implications for DSM-V. *Journal of Abnormal Psychology*, **114**(4), 537–550.
- Lord F.M., Novick M.R. (1968) *Statistical Theories of Mental Test Scores*, Reading, MA, Addison-Wesley.
- Lubke G., Dolan C.V., Kelderman K., Mellenbergh G.J. (2003) On the relationship between sources of within- and between-group differences and measurement invariance in the common factor model. *Intelligence*, **31**(6), 543–566.
- Lubke G., Muthén B. (2005) Investigating population heterogeneity with factor mixture models. *Psychological Methods*, **10**(1), 21–39.
- Lubke G., Muthén B. (2007) Performance of Factor Mixture Models as a function of model size, covariate effects, and class-specific parameters. *Structural Equation Modelling*, **14**(1), 26–47.
- Lubke G., Neale M. (2006) Distinguishing between latent classes and continuous factors: Resolution by Maximum Likelihood? *Multivariate Behavioral Research*, **41**(4), 499–532.
- Markon K.E., Krueger R.F. (2005) Categorical and continuous models of liability to externalizing disorders: A direct comparison in NESARC. *Archives of General Psychiatry*, **62**(12), 1352–1359.
- Mellenbergh G.J. (1989) Item bias and Item Response Theory. *International Journal of Educational Research*, **13**(2), 127–143.
- Meredith W. (1993) Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, **58**(4), 525–543.

- Mineka S., Watson D., Clark L.A. (1998) Comorbidity of anxiety and unipolar mood disorders. *Annual Review of Psychology*, **49**, 377–412.
- Muthén B., Asparouhov T. (2002) *Latent Variable Analysis with Categorical Outcomes: Multiple-Group and Growth Modelling in Mplus*, Mplus Web Notes: No. 4, Los Angeles, CA, Muthén & Muthén.
- Raftery A.E. (1995) Bayesian model selection in social research. *Sociological Methodology*, **25**, 111–163.
- Slade T., Watson D. (2006) The structure of common DSM-IV and ICD-10 mental disorders in the Australian general population. *Psychological Medicine*, **36**(11), 1593–1600.
- Vermunt J.K., Magidson J. (2005) *Technical Guide for Latent GOLD 4.0: Basic and Advanced*, Belmont, MA, Statistical Innovations Inc.
- Vollebergh W.A., Iedema J., Bijl R.V., deGraaf R., Smit F., Ormel J. (2001) The structure and stability of common mental disorders: the NEMESIS study. *Archives of General Psychiatry*, **58**(6), 597–603.
- Watson D. (2005) Rethinking the mood and anxiety disorders: a quantitative hierarchical model for DSM-V. *Journal of Abnormal Psychology*, **114**(4), 522–536.
- Watson D., Clark L.A., Carey G. (1988) Positive and negative affectivity and their relation to anxiety and depressive disorders. *Journal of Abnormal Psychology*, **97**(3), 346–353.
- Wittchen H.-U., Beesdo-Baum K., Gloster A., Höfler M., Klotsche J., Lieb R., Beauducel A., Böhner M., Kessler R.C. (2009) The structure of mental disorders re-examined: Is it developmentally stable and robust against additions? *International Journal of Methods in Psychiatric Research*, **18**(4), 189–203.
- World Health Organization (WHO). (1993) *The ICD-10 Classification of Mental and Behavioural Disorders: Diagnostic Criteria for Research*, Geneva, WHO.