Journal of Affective Disorders 215 (2017) 125–134

Contents lists available at ScienceDirect

Journal of Affective Disorders

journal homepage: www.elsevier.com/locate/jad

Research paper

Using a hybrid subtyping model to capture patterns and dimensionality of depressive and anxiety symptomatology in the general population

Klaas J. Wardenaar[a,*], Rob B.K. Wanders[a], Margreet ten Have[b], Ron de Graaf[b], Peter de Jonge[c]

[a] University of Groningen, University Medical Center Groningen, Interdisciplinary Center Psychopathology and Emotion Regulation (ICPE), Groningen, The Netherlands
[b] Netherlands Institute of Mental Health and Addiction, Utrecht, The Netherlands
[c] University of Groningen, Faculty of Behavioural and Social Sciences, Department of Developmental Psychology, Groningen, The Netherlands

ARTICLE INFO

Keywords:
Depression
Anxiety
Heterogeneity
Subtypes
Latent class analysis
Mixed measurement item response theory

ABSTRACT

Background: Researchers have tried to identify more homogeneous subtypes of major depressive disorder (MDD) with latent class analyses (LCA). However, this approach does no justice to the dimensional nature of psychopathology. In addition, anxiety and functioning-levels have seldom been integrated in subtyping efforts. Therefore, this study used a hybrid discrete-dimensional approach to identify subgroups with shared patterns of depressive and anxiety symptomatology, while accounting for functioning-levels.
Methods: The Comprehensive International Diagnostic Interview (CIDI) 1.1 was used to assess previous-year depressive and anxiety symptoms in the Netherlands Mental Health Survey and Incidence Study-1 (NEMESIS-1; n=5583). The data were analyzed with factor analyses, LCA and hybrid mixed-measurement item response theory (MM-IRT) with and without functioning covariates. Finally, the classes' predictors (measured one year earlier) and outcomes (measured two years later) were investigated.
Results: A 3-class MM-IRT model with functioning covariates best described the data and consisted of a 'healthy class' (74.2%) and two symptomatic classes ('sleep/energy' [13.4%]; 'mood/anhedonia' [12.4%]). Factors including older age, urbanicity, higher severity and presence of 1-year MDD predicted membership of either symptomatic class vs. the healthy class. Both symptomatic classes showed poorer 2-year outcomes (i.e. disorders, poor functioning) than the healthy class. The odds of MDD after two years were especially increased in the mood/anhedonia class.
Limitations: Symptoms were assessed for the past year whereas current functioning was assessed.
Conclusions: Heterogeneity of depression and anxiety symptomatology are optimally captured by a hybrid discrete-dimensional subtyping model. Importantly, accounting for functioning-levels helps to capture clinically relevant interpersonal differences.

1. Introduction

The specific mechanisms underlying depression are still poorly understood, which may partly be due to the heterogeneity of the used categorical depression construct (e.g. Widiger and Clark, 2000). To overcome this problem, researchers have used data-driven statistical models such as Latent Class Analysis (LCA) to identify more homogeneous depression subgroups (Eaton et al., 1989; Kendler et al., 1996; Sullivan and Kendler, 1998; Sullivan et al., 1998, 2002; Parker et al., 1999; Carragher et al., 2009; Hybels et al., 2009; Lamers et al., 2010, 2012; Li et al., 2014; Ulbricht et al., 2015). Promisingly, there is some evidence that the resulting subgroups differentiate persons with distinct treatment responses (Ulbricht et al., 2015), biomarkers and course-trajectories (Lamers et al., 2013, 2016). However, although these studies provide valuable insights into the heterogeneity of depression, the interpretability of LCA results is hampered by the underlying key-assumption that all heterogeneity among persons is explained by class-membership and that no additional variation exists within classes. This means that the discrete LCA models are rather crude approximations of reality, where psychopathology is known to be a continuous phenomenon (Kendell, 1989; Kendell and Jablensky, 2003).

To account for the dimensional nature of psychopathology when identifying data-driven depression subgroups, researchers can use a hybrid mixture approach, such as mixed measurement item response theory (MM-IRT; Rost, 1990, 1991; Mislevy andVerhelst, 1990) models that integrate LCA with an IRT measurement model. MM-IRT is closely related to factor mixture models (FMM, Lubke and

* Corresponding author.

http://dx.doi.org/10.1016/j.jad.2017.03.038
Received 23 November 2016; Received in revised form 9 January 2017; Accepted 10 March 2017
Available online 14 March 2017
0165-0327/

Muthén, 2005; Muthén & Asparoutov, 2006; Lubke and Miller, 2015), which integrate LCA with a factor model (Lubke et al., 2007; Kuo et al., 2008; Picardi et al., 2012; Sunderland et al., 2013; Pattyn et al., 2015; ten Have et al., 2016). In MM-IRT, a measurement model is taken as a point of departure and heterogeneity in response behavior is explained by estimating latent classes for which different IRT-model parameters may hold (e.g. Cohen and Bolt, 2005; Maij-de Meij et al., 2010). Here, LCA and IRT complement each other. On the one hand, one can investigate latent population heterogeneity while accounting for response-behavior variations. On the other hand, one can investigate the dimensionality of symptoms, while accounting for latent population heterogeneity (Clark et al., 2013). Helpfully, covariates can be included in the MM-IRT models to further improve class-differentiation (MM-IRT-C; Tay et al., 2011). MM-IRT has previously been used to investigate response-behavior heterogeneity on personality questionnaires (e.g. Maij-de Meij et al., 2005, 2008; Egberink et al., 2010), patterns of tobacco-use/dependence symptoms (Muthén and Asparouhov, 2006) and the use of special response scales (e.g. Austin et al., 2006).

Apart from not accounting for dimensionality, some other limitations can also be seen in previous subtyping studies. First, many only analyzed depression symptoms, whereas these co-occur more often than not with anxiety symptoms (e.g. Mineka et al., 1998). Indeed, a recent study found that anxiety symptoms may play an important role in the differentiation between depression subgroups (ten Have et al., 2016). Second, previous studies have paid little attention to the role of persons' level of functioning as source of heterogeneity, despite the fact that it is important to determine whether present symptomatology is actually pathological or reflects sub-clinical problems (e.g. Kramer et al., 2004).

Wanders et al. (2016) addressed all the above-described issues by using MM-IRT on data from a large cohort (n=73,403) to identify subgroups based on depressive and anxiety symptoms, while accounting for the role of functioning levels. The results showed that a 5-class MM-IRT-C model with functioning scales incorporated as covariates, optimally differentiated between subgroups with different symptom profiles ('healthy', 'somatic', 'worried', 'subclinical' and 'clinical' subgroups) and patterns of associations with external factors (e.g. sociodemographics, lifestyle). These results clearly showed the promise of MM-IRT-C to identify hybrid discrete-dimensional patient subgroups.

The current study used a similar MM-IRT-C approach to identify cross-diagnostic subtypes of depressive and anxiety symptomatology, incorporating functioning levels as covariates. However, this study also extended on the previous work by making use of a representative population sample (n=5583; the Netherlands Mental Health Survey and Incidence Study-1, NEMESIS-1). In addition, the 3-wave longitudinal design of NEMESIS-I (baseline, 1-year follow-up and 3-year follow-up) allowed for a thorough investigation of the longitudinal correlates of the estimated classes. MM-IRT-C models were estimated on the data collected at 1-year follow-up. Next, the prediction of subgroup membership by baseline variables, as well as the subgroups' prediction of 3-year follow-up outcomes was investigated.

2. Methods

2.1. Participants

Participants came from NEMESIS-1, a longitudinal cohort study in a randomly selected adult population sample (aged 18–65 years) from the Netherlands. The study consisted of a baseline measurement (T0; n=7076; 69.7% response; in the year 1996) a measurement after 1 year (T1; n=6518; 79% response; in the year 1997) and a measurement after 3 years (T2; n=4796; 85% response; in the year 1999). The detailed design, rationale and goals of NEMESIS-1 have been described previously (Bijl et al., 1998). The research protocol was approved by the ethics committee of the Netherlands Institute of Mental Health and Addiction, Utrecht, the Netherlands. All participants provided oral informed consent in line with the prevailing Dutch law at the time the fieldwork took place.

In each measurement-wave, participants were interviewed with the Composite International Diagnostic Interview (CIDI; version 1.1) generating DSM-III-R diagnoses. The depression questions (Section E) and anxiety symptom questions (Section D: Panic Disorder, Generalized Anxiety Disorder [GAD], Agoraphobia, Social Phobia and Specific phobia) were used in the current study. The T1 data (collected 1 year after the baseline measurement) were used to estimate an optimal subtyping model because the time frame of these CIDI symptom assessments was limited to the 1-year period between T0 and T1 (see Supplementary Figure 1). This ensured that the assessed symptoms (co)occurred roughly within the same 1-year time-interval. A previous multivariate analysis showed that sample attrition between T0 and T1 was associated with younger age, lower education, urbanicity, not cohabiting with a steady partner, unemployment, being born outside the Netherlands, agoraphobia, social phobia and eating disorders. The presence of any DSM-III-R disorder was only weakly related to attrition, controlled for demographics (OR=1.20; de Graaf et al., 2000a). Of the 5618 respondents at T1, 5583 (99.4%) provided all the data that was needed for the current analyses (measures of depressive and anxiety symptomatology and functioning). The current analyses were conducted using data from the different measurement waves. The MM-IRT and MM-IRT-C analyses were run in the T1 sample, identifying a range of subgroups. Next, factors assessed at T0 were used to predict subgroup-membership at T1. Finally, the subgroups at T1 were associated with outcomes measured at T2 to evaluate the prognostic value of the identified subgrouping. This was done using all subjects that were included in the MM-IRT analyses and had T2 assessments for the relevant outcomes (see below). When adjusted for sociodemographic factors, attrition between T1 and T2 was associated with the presence of MDD, dysthymia and alcohol dependence (de Graaf et al., 2000b).

2.2. Measurements

2.2.1. Symptom-assessments and functioning at T1

The presence of depressive symptoms was assessed with the depression section of the CIDI 1.1. All depressive symptoms were evaluated irrespective of whether the key-symptoms were endorsed (there were no symptom skips in the depression section). In addition, the responses to CIDI screening questions for a range of common anxiety disorders (Panic Disorder, GAD, Agoraphobia, Social Phobia and Specific phobia) were used in the current study. Here, only the screening questions could be used because the anxiety sections of the CIDI skipped all detailed questions if the screening questions were not endorsed. Taken together, the analyzed symptom-dataset contained 28 depressive symptoms and 5 anxiety symptoms.

The Medical Outcome Study Short Form-36 (SF-36; Stewart et al., 1988) was used to assess several domains of functioning.

2.3. Predictors at T0

Sociodemographic variables (age, gender, employment status, urbanicity and educational attainment) were assessed at baseline. The Mastery scale (Pearlin and Schooler, 1978) was used to assess the extent to which individuals feel in control and/or feel responsibility for the events occurring in their lives. The Rosenberg Self Esteem scale (Rosenberg, 1965) was used to assess self-esteem. The General Health Questionnaire-12 (GHQ-12; Goldberg and Williams, 1988) was used to assess severity and the SF-36 was used to assess levels of functioning. The CIDI was used to determine 1-year CIDI-based DSM-III-R diagnoses, using disorder hierarchies and exclusion rules.

2.4. Outcomes at T2

Five health-related outcomes were constructed based on assessments at T2. The CIDI was used to confirm the presence of a 1-year MDD diagnosis and any 1-year anxiety diagnosis and the SF-36 was used to construct functioning outcomes. Poor psychological functioning, poor social functioning and poor physical functioning were defined as scoring in the lowest tertile of, respectively, the SF-36 psychological health, social functioning and physical functioning scales.

2.5. Statistical analyses

2.5.1. Exploratory factor analyses and latent class analyses

EFA and LCA were conducted first to explore the latent structure of the data. Mean and Variance adjusted Weighted least squares (WLSMV) Exploratory Factor Analysis (EFA) was run on the tetrachoric symptom correlation matrix using Mplus (v 7.0; Muthén and Muthén, 1998–2015). The ratios between the first and subsequent factors' eigenvalues were inspected to evaluate if the data were best described by a one- or multiple-factor model. Next, the goodness-of-fit of the suggested factor model to the whole dataset was investigated by fitting the model as a two-parameter logistic (2PL; Birnbaum, 1968) IRT model with Mplus 7.0, using a WLSMV estimator (this is the default approach to estimate factor models with dichotomous indicator variables in Mplus). The root mean square error of approximation (RMSEA) was used to evaluate the model's absolute fit to the data with an RMSEA≤0.06 indicating adequate fit. LCA was conducted with Latent GOLD (v.5.0; Vermunt and Magidson, 2016) to investigate the heterogeneity in the current sample. Models with increasing numbers of classes were estimated and the Bayesian Information criterion (BIC) and Akaike Information Criterion (AIC) were compared to identify the model that best described the data. Multiple random starts were used to avoid models at local maxima. All model estimations and further analyses were conducted using the appropriate post-stratification weights for the used measurement waves (Bijl et al., 1998).

2.5.2. MM-IRT

In MM-IRT models, a continuous latent dimension (or more than one latent dimension if the EFA shows a multifactorial structure) is modeled within a mixture framework using an IRT model (see Fig. 1). In the used MM-IRT model, the above described 2PL model was used to model the relationship between each dichotomous item and the underlying dimension by its location on the severity dimension (estimated for each item by the threshold parameter) and its discrimination between severity levels (estimated for each item by the slope parameter). More detailed information about the IRT-model parametrization in Latent GOLD can be found in Vermunt and Magidson (2016). In MM-IRT models, the 2PL model is used as a point of departure for the identification of latent subgroupings of individuals with different measurement-model parameters, and thus, different item-response characteristics. Importantly, MM-IRT allows for the identification of subgroups with qualitatively different patterns of response-behavior, while at the same time allowing for dimensional severity variations within classes. This means that persons in the same latent class show similar item-response parameters, but can differ from each other in terms of severity and the number of endorsed symptoms, which can range from only a few mild symptoms to many symptoms, including those at the severe end of the severity spectrum. In the analyses, MM-IRT models based on this IRT model were fit to the data using a maximum likelihood estimator in Latent Gold 5.0 and were compared to each other, using the AIC and BIC. Here, a relative increase in fit of multiple-class models compared to the 1-class MM-IRT model (which is equivalent to the complete sample 2PL model) indicates that item-functioning differs across latent subgroups. The final model was selected based on the lowest BIC and/or AIC.

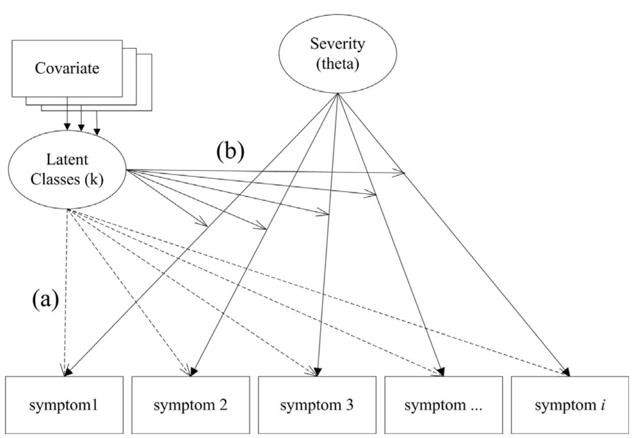

Fig. 1. Schematic representation of Mixed Measurement Item Response Theory with covariates (MM-IRT-C) model: In this model, the probability of symptoms 1 to i depends on the latent trait (theta) and the latent class (k). The discrimination (b) and location (a) parameters are moderated by class membership. Also, the effects of covariates on class membership are modeled.

After estimation of the regular MM-IRT models, the MM-IRT-C models were run using the same estimation methods, but adding the SF-36 physical functioning, social functioning, physical role functioning, and emotional role functioning scales as covariates (see also Fig. 1). These four scales were included because they were expected to add unique information about inter-personal variations in health-related functioning, on top of the information provided by the symptom data. Both role-functioning scales were dichotomized (100%=1 and <100%=0). First, the social and physical functioning scales were added. Second, the physical and emotional role-functioning scales were added. Finally, all four scales were added as covariates.

2.6. External associations

After identification of the best model, its associations with external variables were investigated. First, associations between variables at T0 and class membership at T1 were investigated. Univariate and multivariate multinomial regression analyses were run, using the categorical latent class variable as outcome and sociodemographic, psychiatric and psychological variables at T0 as independent variables. Second, class membership at T1 was used to predict the five constructed outcomes at T2. Logistic regression models were run with each of these outcomes as dependent variable and the latent class dummy-variables as independent variables. Unadjusted models were run first, followed by models adjusted for T1 demographic variables (i.e. gender, age, education, urbanicity, living situation and employment). Finally, models were rerun adjusting for both the GHQ-12 score and the outcome's value at T1.

3. Results

3.1. Baseline descriptive information

Of the selected 5583 participants, 49.7% was female and the mean age was 39.3 years. Of the sample, 64.6% was employed, 15.7% was homemaker, 7.1% was student, 6.2% was unemployed or disabled and 6.3% was retired. Of the participants, 82.5% percent came from an urban area, 69.8% lived together with a spouse and 30.1% had a college education. In the year prior to baseline, 5.6% had MDD, 2.3% had panic disorder, 1.0% had GAD, 1.2% had agoraphobia without panic disorder, 4.4% had social phobia, and 6.9% had specific phobia.

3.2. Exploratory factor analysis and latent class analysis

The Eigenvalues of the first five factors were 18.4, 1.7, 1.6, 1.3 and 1.1, respectively. The ratio between the first and second factors'

Table 1
Results of fitting latent class models and MM-IRT models to the depression and anxiety symptom data at T1.

Model	Covariates	Model	Log likelihood	Number of parameters	BIC	AIC
LCA	–	2-class	−33564.8	67	67707.6	67263.5
		3-class	−31980.4	101	64832.3	64162.9
		4-class	−31614.3	135	64393.4	63498.6
		5-class	−31382.6	169	64223.4	63103.2
		6-class	−31223.0	203	<u>64197.4</u>	62851.9
		7-class	−31089.5	237	64223.9	62653.1
MM-IRT	–	1-class	−31946.6	66	64462.7	64025.3
		2-class	−31251.8	133	63651.0	62769.5
		3-class	−30921.9	200	63569.3	62243.7
		4-class	−30731.0	267	63765.6	61995.9
MM-IRT-C	PF & SF	2-class	−30943.1	135	63050.9	62156.1
		3-class	−30608.4	204	62977.0	61624.9
		4-class	−30352.7	273	63060.8	61251.4
	PRF & ERF	2-class	−30951.2	135	63067.2	62172.5
		3-class	−30626.8	204	63013.7	61661.6
		4-class	−30392.9	273	63141.3	61331.8
	PF, SF, PRF & ERF	2-class	−30844.2	137	62870.5	61962.5
		3-class	−30512.8	208	<u>62820.3</u>	61441.7
		4-class	−30270.3	279	62947.8	61098.6

All analyses were weighted. LCA=Latent class analysis; MM-IRT=Mixed Measurement Item response theory; MM-IRT-C=MM-IRT with covariate(s); BIC=Bayesian Information Criterion; AIC=Akaike Information Criterion; PF=SF36 physical functioning scale; SF=SF-36 social functioning scale; PRF=physical role functioning; ERF=emotional role functioning.

Eigenvalues was 11.5, indicating that the first factor described the majority of the variance in the data. In the 1-factor model, all items showed substantial factor loadings (range: 0.44–0.89). In addition, a fitted unidimensional IRT model (2PL) showed good fit to the data (RMSEA=0.024 [90%CI =0.023–0.025). Based on these results, it was decided to model a single latent dimension in all MM-IRT models.

The LCA results (Table 1) showed a decreasing BIC for models with up to 6 classes. Although the AIC did decrease with further class-additions, the 6-class model was further inspected. The model consisted of a 'healthy' (69.7), 'sleep problems' (6.7%), 'lack of energy' (13.7%), 'moderate somatic depression' (5.1%), 'moderate cognitive depression' (2.8%) and 'severe' (2.0%) class (see Supplementary Figure 2).

3.3. Mixture-IRT

The MM-IRT models had lower BIC and AIC values than the LCA models (Table 1), indicating that the hybrid MM-IRT models described the data better. Also, MM-IRT models with two or more classes had lower BIC/AIC values than the single-class MM-IRT model, indicating that multiple classes with different IRT parameters better described the data than a single 2PL model for the whole sample.

MM-IRT-C models described the data better than regular MM-IRT models (lower AIC/BIC). The 3-class model with all four functioning scales added as covariates described the data best (lowest BIC; entropy=0.76). The classes of this model were characterized by different symptom-endorsement patterns (Fig. 2). The first class was characterized by low endorsement of all symptoms (healthy: 74.2%). The second class (sleep/energy: 13.4%) was characterized by increased endorsement of sleeping problems, morning tiredness and energy loss. The third class (mood/anhedonia: 12.4%) also showed increased endorsement of sleep problems and energy loss, but also showed high endorsement of loss of interest, anhedonia and concentration/decision-making problems.

SF-36 social functioning (Wald statistic=179.6; $p < 0.001$), physical functioning (Wald statistic=10.4; $p=0.006$), physical role functioning (Wald statistic=8.0; $p=0.02$), and emotional role functioning (Wald statistic=185.9; $p < 0.001$) were all significantly associated with class membership. Inspection of the posterior probabilities of class-membership for different levels of functioning (Supplementary Table 1) showed that scoring high on each of the scales was associated with a higher probability of healthy-class membership. Lower scores were associated with higher probabilities of being classified in one of the two symptomatic classes.

3.4. Class-membership prediction

In analyses to investigate the predictive associations of T0 predictors with T1 class membership, both symptomatic classes were first compared to the healthy class (Table 2) and next compared to each other (Table 3). In univariate analyses almost all T0 predictors were significantly associated with increased odds of being in the sleep/energy class or mood/anhedonia class compared to the healthy class. In multivariate analyses, several of these effects were no longer significant. Higher odds of being in the mood/anhedonia class rather than the healthy class were associated with female gender (OR=1.48), urban environment (OR=1.35), paid employment (OR=1.34), higher GHQ-12 (OR=1.11), 1-year MDD (OR=3.12), 1-year panic disorder (OR=1.85), 1-year GAD (OR=2.11), 1-year agoraphobia (OR=2.04), 1-year specific phobia (OR=1.65), lower SF-36 social functioning (OR=0.92), lower SF-36 vitality (OR=0.91), lower SF-36 psychological health (OR=0.81), lower SF-36 general health (OR=0.93) and lower mastery (OR=0.96) at T0. Higher odds of being in the sleep/energy class rather than the healthy class were associated with female gender (OR=1.52), urban environment (OR=1.46), living alone (OR=1.38), higher GHQ-12 (OR=1.08), 1-year MDD (OR=3.19), 1-year panic disorder (OR=2.14), 1-year agoraphobia (OR=2.26), 1-year specific phobia (OR=1.46), lower SF-36 psychological health (OR=0.87), lower SF-36 general health (OR=0.93) and lower self-esteem (OR=0.95) at T0. Univariate comparisons of the symptomatic classes using 'sleep/energy' as reference (Table 3), showed that a lower age (OR=0.95), paid employment (OR=1.36), higher GHQ-12 (OR=1.07), lower SF-36 social functioning (OR=0.94), lower SF-36 vitality (OR=0.93) and lower SF-36 psychological health (OR=0.91) at T0 were significantly associated with a higher odds of being in the mood/anhedonia class than in the sleep/energy class. In the multivariate analyses only a lower age (OR=0.95) and paid employment (OR=1.40) retained significant predictive effects.

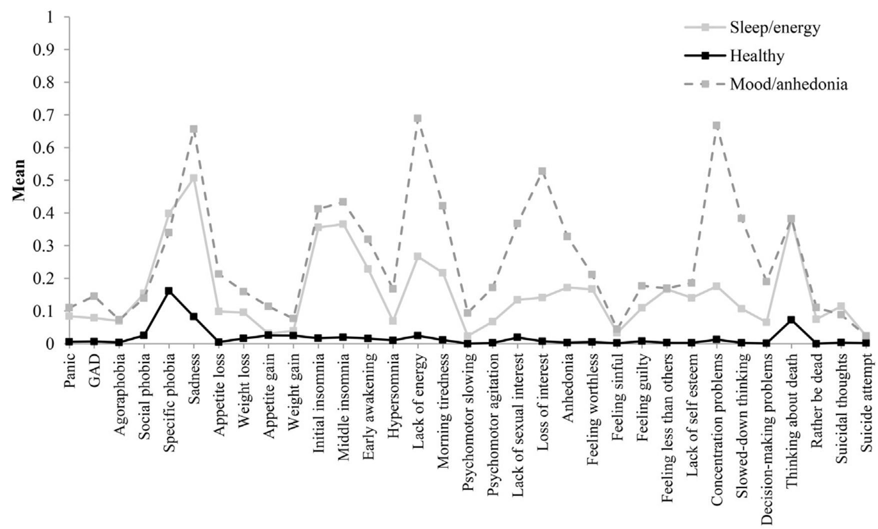

Fig. 2. Mean item scores in each of the estimated classes of the 3-class MM-IRT-C model.

3.5. Prediction of T2 outcomes

In unadjusted prediction models (Table 4), membership of the ‘sleep/energy’ class and the ‘mood/anhedonia’ class both predicted higher odds of a MDD diagnosis, any anxiety diagnosis, poor psychological health, poor social functioning and poor physical functioning at T2. Effects decreased but mostly remained significant when adjusted for T1 covariates. Comparisons between the two symptomatic classes showed that only the odds of MDD at T2 were higher in the ‘mood/anhedonia’ class than in the ‘sleep/energy’ class (OR=1.79). Other associations were non-significant.

4. Discussion

This study aimed to identify hybrid discrete-dimensional subgroups to optimally capture heterogeneity in depressive and anxiety symptomatology in a population sample. The results showed that the data were best described by a 3-class MM-IRT-C model with functioning scales added as covariates. This model consisted of a healthy, sleep/energy and mood/anhedonia class. Several factors measured at T0, including female gender, GHQ-12 score, 1-year MDD, the presence of several 1-year anxiety disorders, and several functioning domains were all predictive of being in either of the symptomatic classes rather than the healthy class at T1. However, only age and employment status at T0 were found to differentiate between the sleep/energy and mood/anhedonia classes at T1. Interestingly, both being in the sleep/energy class and being in the mood/anhedonia class was associated with poorer outcomes at T2 compared to the healthy class. Also, being in the mood/anhedonia class was predictive of higher odds of 1-year MDD at follow-up, compared to the sleep/energy class. No such differentiation was observed on any of the other outcomes. These results indicated that MM-IRT-C identified subgroups that clearly differed in terms of their symptom-patterns and external correlates, although predictors mainly differed between the healthy class and the symptomatic classes, and less between the sleep/energy and mood/anhedonia classes.

The fact that a MM-IRT model fit the data better than LCA models showed the importance of accounting for dimensionality when aiming to optimally subtype persons based on their symptoms. The finding that a MM-IRT-C model with 3 classes fit the data better than the best-fitting LCA model with 6 classes is in line with previous notions that hybrid models can provide solutions with fewer classes than LCA models, especially when the symptoms have a strong underlying (uni) dimensional structure (Clark et al., 2013). The findings also align with the previous findings by Wanders et al. (2016), although that study identified more classes, including a purely somatic class, a subclinical class and more specific symptomatic classes. The current findings can also be compared with findings from closely-related FMM studies. In a large sample of treatment-seeking patients (n=1165), Sunderland et al. (2013) found that hybrid FMMs better described depressive symptom heterogeneity than LCA models, as indicated by lower AIC and BIC values, although the authors concluded that a more parsimonious factor model outperformed the FMMs. Conversely, in a symptomatic population sample (NEMESIS-II; n=1388), ten Have et al. (2016) found that a LCA with a freely estimated within-class correlation between appetite gain and appetite-loss, described depression data better than a range of tested hybrid FMMs. Variations in the findings across the referenced studies and the current study could be due to differences in sample composition (more healthy vs. more symptomatic; select vs. random samples), the time-interval for which symptoms were assessed (past two weeks [Sunderland et al., 2013] vs. worst lifetime episode [ten Have et al., 2016]), the used assessment instruments (Patient Health Questionnaire [PHQ-9; Sunderland et al., 2013] vs. CIDI [ten Have et al., 2016] vs. Mini International Neuropsychiatric Interview [MINI; Wanders et al., 2016]) and the exact analytical approach. For instance, ten Have et al. (2016) used a regular LCA and an LCA-variant that allows for some local dependence within the model, whereas Sunderland et al. (2013) and the current study only employed regular LCA.

An interesting observation was that a model with SF-36 (role-) functioning scales added as covariates provided a better description of the data. This aligns with the results presented by Wanders et al. (2016) and shows that functioning/disability levels should be considered an important source of heterogeneity, in line with existing ideas about the relevance of disability as an indicator of pathological problems and need for care (Kramer et al., 2004; McKnight and Kashdan, 2009). More generally, these findings show that the incorporation of more available information in a data-driven subtyping model helps to improve the explanation of sample heterogeneity. Remarkably, this has so far received little attention in the depression subtyping literature. Previous studies have focused on identification of purely symptom-based subgroups, only investigating associations with other variables in post-hoc analyses to evaluate the subgroups’ validity (e.g. Kendler et al., 1996; Lamers et al., 2010, ten Have et al., 2016). The current results suggest that it might be preferable to already include other relevant information (i.e. functioning, medication-use, demographics etc.) in the actual subtyping models. In that way, more sources of inter-individual variation are incorporated in and explained by the model, yielding subtypes that are not only better in terms of

Table 2
Prediction of class membership at T1 by clinical and psychological predictors measured one year earlier at T0.

	Class 1: Healthy (n=4236)		*Class 2:* Mood/Anhedonia (n=656)			*Class 3:* Sleep/energy (n=694)		
				Univariate	Multivariate		Univariate	Multivariate
	Descriptives		Descriptives	OR (95%CI)	OR (95%CI)	Descriptives	OR (95%CI)	OR (95%CI)
Female gender, n (%)	1924 (45.4%)	ref.	393 (59.9%)	1.80 (1.52–2.13)***	1.48 (1.22–1.79)***	429 (61.8%)	1.95 (1.65–2.29)***	1.52 (1.26–1.83)***
Age, mean (s.d.)[b]	39.0 (12.7)	ref.	39.0 (11.3)	1.00 (0.97–1.04)	1.00 (0.96–1.04)	40.5 (12.6)	1.05 (1.02–1.09)**	1.03 (0.99–1.07)
Urban environment, n (%)	3448 (81.4%)	ref.	568 (86.7%)	1.49 (1.17–1.89)**	1.35 (1.04–1.74)*	605 (87.1%)	1.55 (1.22–1.96)***	1.46 (1.13–1.88)**
College education, n (%)	1255 (29.6%)	ref.	200 (30.4%)	1.04 (0.87–1.24)	–	212 (30.6%)	1.05 (0.88–1.24)	–
Living alone, n (%)	666 (15.7%)	ref.	143 (21.8%)	1.49 (1.22–1.83)***	1.12 (0.87–1.40)	167 (24.0%)	1.70 (1.40–2.06)***	1.38 (1.11–1.72)**
Paid employment, n (%)	2785 (65.8%)	ref.	409 (62.4%)	0.86 (0.73–1.02)	1.34 (1.09–1.64)**	381 (55.0%)	0.64 (0.54–0.75)***	1.04 (0.86–1.26)
GHQ-12 score, median (IQR)	0.0 (0.0–1.0)	ref.	1.0 (0.0–4.0)	1.41 (1.37–1.46)***	1.11 (1.06–1.17)***	1.0 (0.0–3.0)	1.33 (1.28–1.38)***	1.08 (1.03–1.14)**
1-year MDD, n (%)	85 (2.0%)	ref.	115 (17.5%)	10.35 (7.71–13.9)***	3.12 (2.20–4.41)***	110 (15.8%)	9.17 (6.82–12.33)***	3.19 (2.26–4.50)***
1-year Panic Disorder, n (%)	34 (0.8%)	ref.	44 (6.8%)	9.09 (5.76–14.34)***	1.85 (1.05–3.27)*	48 (6.9%)	9.22 (5.88–14.44)***	2.14 (1.23–3.71)**
1-year GAD, n (%)	18 (0.4%)	ref.	22 (3.4%)	8.21 (4.39–15.34)***	2.11 (1.01–4.42)*	15 (2.1%)	5.08 (2.55–10.15)***	1.42 (0.64–3.16)
1-year Agoraphobia, n (%)	23 (0.5%)	ref.	21 (3.2%)	6.17 (3.40–11.20)***	2.04 (1.02–4.07)*	24 (3.4%)	6.49 (3.63–11.60)***	2.26 (1.17–4.38)*
1-year Social phobia, n (%)	103 (2.4%)	ref.	66 (10.0%)	4.47 (3.24–6.16)***	1.21 (0.80–1.82)	76 (10.9%)	4.93 (3.62–6.71)***	1.45 (0.99–2.14)
1-year Specific phobia, n (%)	189 (4.5%)	ref.	101 (15.4%)	3.90 (3.02–5.05)***	1.65 (1.20–2.27)**	93 (13.4%)	3.31 (2.55–4.31)***	1.46 (1.06–2.01)*
Functioning scales, median (IQR)								
SF-36 Physical functioning[a]	100 (95–100)	ref.	95 (85–100)	0.81 (0.78–0.85)***	1.01 (0.94–1.08)	95 (85–100)	0.79 (0.75–0.82)***	0.96 (0.90–1.03)
SF-36 Social functioning[a]	100 (90–100)	ref.	87.5 (67.5–100)	0.68 (0.65–0.71)***	0.92 (0.86–0.97)**	90 (67.5–100)	0.72 (0.69–0.75)***	0.97 (0.91–1.03)
SF-36 Vitality[a]	75 (65–85)	ref.	60 (45–75)	0.64 (0.61–0.67)***	0.91 (0.85–0.98)*	65 (50–80)	0.68 (0.65–0.71)***	0.95 (0.89–1.02)
SF-36 Pain[a]	100 (80–100)	ref.	90 (67.5–100)	0.83 (0.81–0.86)***	0.98 (0.93–1.03)	90 (67.5–100)	0.82 (0.80–0.85)***	0.96 (0.92–1.01)
SF-36 Psychological health[a]	88 (80–92)	ref.	76 (60–84)	0.54 (0.51–0.57)***	0.81 (0.74–0.88)***	76 (60–88)	0.59 (0.56–0.62)***	0.87 (0.80–0.95)**
SF-36 General health[a]	80 (70–90)	ref.	70 (55–85)	0.73 (0.70–0.76)***	0.93 (0.87–0.99)*	70 (55–85)	0.73 (0.70–0.76)***	0.93 (0.87–0.98)*
Mastery scale, mean (s.d.)	20.0 (3.0)	ref.	17.8 (3.9)	0.83 (0.81–0.85)***	0.96 (0.93–1.00)*	17.9 (3.7)	0.83 (0.81–0.85)***	0.97 (0.94–1.01)
Rosenberg Self-esteem scale, mean (s.d.)	33.6 (3.8)	ref.	31.2 (4.8)	0.87 (0.85–0.88)***	0.99 (0.97–1.02)	31.0 (4.7)	0.86 (0.84–0.87)***	0.95 (0.93–0.98)***

OR=odds ratio; 95%CI=95% Confidence Interval; GHQ-12=General Health Questionnaire-12; MDD= Major Depressive Disorder; GAD=Generalized Anxiety Disorder, MOS-SF 36=Medical Outcome Study Short Form-36.

[a] Odds ratios are given for 10-point increments.

[b] Odds ratios given for 5-year increments

* $p < 0.05$,

** $p < 0.01$;

*** $p < 0.001$.

Table 3
Prediction of membership of the 'sleep/energy' vs. 'mood/anhedonia' classes.

	T1 class-membership		
	Sleep/energy (n=694)	Mood/Anhedonia (n=656)	
		Univariate[a]	Multivariate[a]
T0 predictors		OR (95%CI)	OR (95%CI)
Female gender, n (%)	Ref	0.93 (0.74–1.15)	–
Age, mean (s.d.)	Ref	0.95 (0.91–0.99)*	0.95 (0.91–1.00)*
Urban environment, n (%)	Ref	0.96 (0.70–1.32)	–
College education, n (%)	Ref	1.05 (0.94–1.17)	–
Living alone, n (%)	Ref	0.88 (0.68–1.13)	–
Paid employment, n (%)	Ref	1.36 (1.09–1.69)**	1.40 (1.12–1.76)**
GHQ-12 score, median (IQR)	Ref	1.07 (1.03–1.11)***	1.04 (0.99–1.10)
1-year MDD, n (%)	Ref	1.13 (0.85–1.50)	–
1-year Panic Disorder, n (%)	Ref	0.99 (0.65–1.51)	–
1-year GAD, n (%)	Ref	1.62 (0.83–3.14)	–
1-year Agoraphobia, n (%)	Ref	0.95 (0.52–1.73)	–
1-year Social phobia, n (%)	Ref	0.91 (0.64–1.28)	–
1-year Specific phobia, n (%)	Ref	1.18 (0.87–1.60)	–
Functioning scales, median (IQR)			
SF-36 Physical functioning	Ref	1.04 (0.98–1.10)	–
SF-36 Social functioning	Ref	0.94 (0.90–0.99)**	0.97 (0.91–1.03)
SF-36 Vitality	Ref	0.93 (0.89–0.98)*	0.99 (0.92–1.08)
SF-36 Pain	Ref	1.01 (0.97–1.06)	–
SF-36 Psychological health	Ref	0.91 (0.86–0.97)**	0.96 (0.87–1.05)
SF-36 General health	Ref	1.00 (0.94–1.05)	–
Mastery scale, mean (s.d.)	Ref	0.99 (0.96–1.02)	–
Rosenberg Self-esteem scale, mean (s.d.)	Ref	1.01 (0.99–1.03)	–

GHQ-12=General Health Questionnaire-12; MDD= Major Depressive Disorder; GAD=Generalized Anxiety Disorder, SF-36=Medical Outcome Study Short Form-36; IQR=interquartile range.
[b]OR given per 5-year increase.
[c]OR given per 10-point increase.
[a] Logistic regression analyses using class-membership (0=sleep/energy and 1=mood/anhedonia) as outcome.
* $p < 0.05$;
** $p < 0.01$;
*** $p < 0.001$.

statistical fit, but also in terms of differentiation between clinical pictures. Although mixture models can become very complex when multiple sources of variation are included, including the additional variables as covariates appears like a suitable way to incorporate more information into the model at the expense of only a few added model parameters.

The current finding of a 'sleep/energy' and 'mood/anhedonia' class aligns with observations from previous data-driven studies of depression heterogeneity. Several factor-analytical studies have shown that depression symptomatology can be decomposed into a mood/cognition-related and a somatic symptom factor (e.g. Shafer, 2006; Wardenaar et al., 2010), which have been shown to be differently related with prospective outcomes, with mood-cognitive symptoms being specifically predictive of MDD and somatic symptoms being specifically predictive of anxiety at 2-year follow-up (Wardenaar et al., 2012). The current findings also showed that being in the mood/anhedonia class was specifically predictive of MDD at follow-up. However, anxiety at follow-up did not differ significantly between the classes, possibly due to the fact that both classes showed relatively similar levels of sleep and anxiety symptoms. Other studies have shown distinct roles of mood/cognitive and somatic symptoms in the course of depression. For instance, Monden et al. (2016) reported that, longitudinal depression data could be decomposed into those showing mainly persisting cognitive/affective symptoms, those showing persisting somatic symptoms and those with quick overall recovery. Wardenaar et al. (2015) showed that patients could be subdivided into classes with different patterns of course-trajectories on somatic vs. mood/cognitive symptomatology. Although methodologically different, all of the abovementioned work indicates that a considerable part of the heterogeneity in depression is related to the distinction between mood/cognitive and somatic/vegetative symptoms.

The results showed no clear distinction between pure depression and anxiety cases, which is in line with previous MM-IRT-C results (Wanders et al., 2016). Also, these findings fit in the larger literature on depression and anxiety comorbidity, showing that the disorder groups more often occur together than alone and very likely share common underlying risk-factors and etiological mechanisms (e.g. Mineka et al., 1998).

Interestingly, unlike previous LCA studies that have found atypical and typical subtypes, characterized by appetite/weight-gain and appetite/weight-loss, respectively (e.g. Lamers et al., 2010), the current classes did not show such a distinction. This is likely to be due to the different ways in which local dependence between appetite/weight gain and loss (with a correlation close to |1|) influences LCA and MM-IRT models. Local dependence can lead to the identification of method-based classes in LCA, classifying those reporting appetite/weight-loss and those reporting appetite/weight-gain to distinct classes to neutralize the local dependence (van Loo et al., in press). This issue was recently illustrated by ten Have et al. (2016), who showed that accounting for the local dependence between appetite/weight loss and appetite/weight gain in a LCA model led to a model that statistically better described the data, but no longer had distinct appetite/weight loss and gain classes. In MM-IRT and FMM, (co) variation within classes is allowed and modeled with a latent dimensional model. This might weaken the effect of local dependencies and prevent the differentiation between classes based on locally dependent symptoms alone.

Although this study had several strengths (e.g. large sample size,

Table 4
Predictive associations between class-membership at T1 with clinical outcomes two years later at T2.

	T2 Outcomes									
	1-year		1-year		Lowest tertile: SF-36		Lowest tertile: SF-36		Lowest tertile: SF-36	
	MDD diagnosis		Anxiety diagnosis		Psychological health		Social functioning		Physical functioning	
Prediction models	n	OR (95% CI)	n	OR (95% CI)	n	OR (95%CI)	n	OR (95%CI)	n	OR (95%CI)
Crude										
Mood/anhedonia vs. rest	4768	7.32 (5.34–10.04)***	4768	5.38 (3.99–7.27)***	4768	3.33 (2.77–4.00)***	4768	2.88 (2.39–3.47)***	4768	2.03 (1.69–2.44)***
Sleep/energy vs. rest	4768	3.54 (2.45–5.13)***	4768	5.25 (6.90–7.08)***	4768	3.04 (2.54–3.64)***	4768	2.77 (2.30–3.32)***	4768	2.19 (1.83–2.62)***
Mood/anhedonia vs. sleep/energy	1161	2.07 (1.41–3.03)***	1161	1.03 (0.74–1.43)	1161	1.09 (0.87–1.38)	1161	1.04 (0.82–1.32)	1161	0.93 (0.73–1.18)
Adjusted 1										
Mood/anhedonia vs. rest	4708	6.85 (4.97–9.45)***	4708	5.10 (3.74–6.95)***	4708	3.12 (2.58–3.76)***	4708	2.67 (2.21–3.23)***	4708	2.01 (1.65–2.44)***
Sleep/energy vs. rest	4708	3.15 (2.15–4.63)***	4708	5.06 (3.71–6.90)***	4708	2.74 (2.28–3.30)***	4708	2.42 (2.00–2.92)***	4708	1.96 (1.62–2.38)***
Mood/anhedonia vs. sleep/energy	1144	2.23 (1.50–3.30)***	1144	1.03 (0.73–1.45)	1144	1.15 (0.90–1.46)	1144	1.11 (0.87–1.41)	1144	1.02 (0.80–1.31)
Adjusted 2										
Mood/anhedonia vs. rest	4702	4.30 (2.89–6.40)***	4702	2.37 (1.63–3.47)***	4698	1.19 (0.94–1.49)	4702	1.47 (1.18–1.84)**	4702	1.54 (1.22–1.94)***
Sleep/energy vs. rest	4702	2.65 (1.78–3.95)***	4702	2.80 (1.98–3.96)***	4698	1.41 (1.14–1.74)**	4702	1.67 (1.36–2.05)***	4702	1.43 (1.15–1.78)**
Mood/anhedonia vs. sleep/energy	1141	1.79 (1.17–2.72)**	1141	0.89 (0.61–1.29)	1139	0.92 (0.71–1.19)	1141	0.95 (0.73–1.23)	1141	1.09 (0.82–1.44)

OR=Odds ratio; 95%CI=95% Confidence Interval. SF-36=Medical Outcome Study Short Form-36.
The results are based on multinomial regression analyses using the healthy class as reference category.
Adjusted 1: T1 gender, age, urbanicity, living alone, employment status.
Adjusted 2: T1 gender, age, urbanicity, living alone, employment status, GHQ-12 score and status on the outcome at T1.

CIDI data without any skips, availability of predictors and outcomes), there were also some study limitations. First, the results apply to part of a population sample and cannot be directly generalized to other (clinical) samples. Second, data were collected with an older version of the CIDI and DSM. However, for this study's purposes, this limitation was outweighed by the fact that full CIDI depressive symptom-assessments were available without skips and the fact that the latent structure of depressive and anxiety symptomatology is likely to have remained stable. Third, the time frame for the assessed symptoms was one year. This meant that reported symptoms did not necessarily all occur at the same time, although by using the T1 data, it was at least made sure that the interval between reported symptoms never exceeded one year. Fourth, the SF-36 scales were assessed for T1, whereas the assessed symptoms could have occurred anywhere between T0 and T1. Fifth, the sample size was not large enough to allow for a random split into a model-development sample and a model-validation sample for confirmatory modeling. In future studies, the current results should be replicated in independently collected data. In addition, the explored models could be extended with a broader range of symptoms and more or other covariates.

There are also some limitations to the analytical approach specifically that should be kept in mind. Although the use and usefulness of MM-IRT-C models have been investigated previously (e.g. Tay et al., 2011; Maij-de Meij et al., 2008), optimal selection strategy for the number of classes under different modeling conditions (e.g. different measurement models; different sample sizes) is still relatively under studied, especially for the MM-IRT variant with covariates (Tay et al., 2011). Future (simulation) studies could provide more insights into this. Another concern is the fact that after model estimation, persons were allocated to a latent class based on their highest posterior class probability, as is often done in mixture analyses. However, this can lead to a level of uncertainty in the class allocation. Some persons may show a clear highest probability for one particular class, making it easy to allocate them to a class, whereas other persons' patterns of class-probabilities may indicate roughly equal probabilities for two or more classes, making it harder to allocate these persons to a class with high certainty. Possibilities to account for this uncertainty could be investigated in future studies.

In conclusion, the heterogeneity in depression and anxiety symptomatology was best described by a hybrid discrete-dimensional subtyping model. Moreover, incorporating information about persons' functioning was shown to lead to a subtyping model that even better described interpersonal heterogeneity. Apart from these findings' implications for depression and anxiety subtyping, the current study should mainly be seen as a proof-of-principle for (1) the use of MM-IRT in psychopathology subtyping and (2) the use of clinical covariates in data-driven subtyping models. Ultimately, the use of these flexible analytical approaches could contribute to the identification of psychopathology subtypes that combine adequate fit to empirical data with optimal usefulness in research and clinical settings.

Appendix A. Supplementary material

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.jad.2017.03.038.

References

Austin, E.J., Deary, I.J., Egan, V., 2006. Individual differences in response scale use: mixed Rasch modelling of responses to NEO–FFI items. Pers. Individ. Dif. 40, 1235–1245.

Bijl, R.V., van Zessen, G., Ravelli, A., de Rijk, C., Langendoen, Y., 1998. The Netherlands Mental Health Survey and Incidence Study (NEMESIS): objectives and design. Soc. Psych. Psych. Epid. 33 (12), 581–586.

Birnbaum, A., 1968. Some latent traits models and their use in inferring an examinee's ability. In: Lord, F.M., Novick, M.R. (Eds.), Statistical Theories of Mental Test Scores. Addison-Wesley, Reading, 397–479.

Carragher, N., Adamson, G., Bunting, B., McCann, S., 2009. Subtypes of depression in a nationally representative sample. J. Affect. Disord. 113, 88–99.

Clark, S.L., Muthén, B., Kaprio, J., D'Onofrio, B.M., Viken, R., Rose, R.J., Smalley, S.L., 2013. Models and strategies for factor mixture analysis: two examples concerning the structure underlying psychological disorders. Struct. Equ. Model. 20 (4), 681–703.

Cohen, A.S., Bolt, D.M., 2005. A mixture model analysis of differential item functioning. J. Educ. Meas. 42, 133–148.

de Graaf, R., Bijl, R.V., Smit, F., Ravelli, A., Vollebergh, W.A., 2000a. Psychiatric and sociodemographic predictors of attrition in a longitudinal study: The Netherlands Mental Health Survey and Incidence Study (NEMESIS). Am. J. Epidemiol. 152 (11), 1039–1047.

de Graaf, R., Bijl, R.V., Vollebergh, W.A.M., van Dorsselaer, S., Spijker, J., Smit, F., Laitinen-Krispijn, S., 2000b. Response and Non-Response Third Wave: the Netherlands Mental Health Survey and Incidence Study (NEMESIS) Technical Report no. 11. Trimbos-Institute, Utrecht.

Eaton, W.W., Dryman, A., Sorenson, A., McCutcheon, A., 1989. DSM-III major depressive disorder in the community. A latent class analysis of data from the NIMH epidemiologic catchment area programme. Br. J. Psychiatry 155, 48–54.

Egberink, I.J.L., Meijer, R.R., Veldkamp, B.P., 2010. Conscientiousness in the workplace: applying mixture IRT to investigate scalability and predictive validity. J. Res. Personal. 44, 232–244.

Goldberg, D.P., Williams, P., 1988. A user's guide to the General Health Questionnaire. NFER-Nelson, Windsor, UK.

Hybels, C.F., Blazer, D.G., Pieper, C.F., Landerman, L.R., Steffens, D.C., 2009. Profiles of depressive symptoms in older adults diagnosed with major depression: latent cluster analysis. Am. J. Geriatr. Psychiatry 17 (5), 387–396.

Kendell, R., Jablensky, A., 2003. Distinguishing between the validity and utility of psychiatric diagnoses. Am. J. Psychiatry 160, 4–12.

Kendell, R.E., 1989. Clinical validity. Psychol. Med. 19, 45.

Kendler, K.S., Eaves, L.J., Walters, E.E., Neale, M.C., Heath, A.C., Kessler, R.C., 1996. The identification and validation of distinct depressive syndromes in a population-based sample of female twins. Arch. Gen. Psychiatry 53 (5), 391–399.

Kramer, T.L., Smith, G.R., Maruish, M.E., 2004. Major depressive disorder: assessing the spectrum of symptoms and functional impairment at treatment entry and follow-up. In: Maruish, M.E. (Ed.), The Use of Psychological Testing for Treatment Planning and Outcomes Assessment, Instruments for Adults Vol. 33rd edition. Lawrence Erlbaum Associates Publishers, Mahwah, NJ, 293–311.

Kuo, P.H., Aggen, S.H., Prescott, C.A., Kendler, K.S., Neale, M.C., 2008. Using a factor mixture modeling approach in alcohol dependence in a general population sample. Drug Alcohol Depend. 98 (1–2), 105–114.

Lamers, F., de Jonge, P., Nolen, W.A., Smit, J.H., Zitman, F.G., Beekman, A.T., Penninx, B.W., 2010. Identifying depressive subtypes in a large cohort study: results from the Netherlands Study of Depression and Anxiety (NESDA). J. Clin. Psychiatry 71 (12), 1582–1589.

Lamers, F., Burstein, M., He, J.P., Avenevoli, S., Angst, J., Merikangas, K.R., 2012. Structure of major depressive disorder in adolescents and adults in the US general population. Br. J. Psychiatry 201, 143–150.

Lamers, F., Vogelzangs, N., Merikangas, K.R., de Jonge, P., Beekman, A.T., Penninx, B.W., 2013. Evidence for a differential role of HPA-axis function, inflammation and metabolic syndrome in melancholic versus atypical depression. Mol. Psychiatry 18 (6), 692–699.

Lamers, F., Beekman, A.T., van Hemert, A.M., Schoevers, R.A., Penninx, B.W., 2016. Six-year longitudinal course and outcomes of subtypes of depression. Br. J. Psychiatry 208 (1), 62–68.

Li, Y., Aggen, S., Shi, S., Gao, J., Li, Y., Tao, M., Zhang, K., Wang, X., Gao, C., Yang, L., Liu, Y., Li, K., Shi, J., Wang, G., Liu, L., Zhang, J., Du, B., Jiang, G., Shen, J., Zhang, Z., Liang, W., Sun, J., Hu, J., Liu, T., Wang, X., Miao, G., Meng, H., Li, Y., Hu, C., Li, Y., Huang, G., Li, G., Ha, B., Deng, H., Mei, Q., Zhong, H., Gao, S., Sang, H., Zhang, Y., Fang, X., Yu, F., Yang, D., Liu, T., Chen, Y., Hong, X., Wu, W., Chen, G., Cai, M., Song, Y., Pan, J., Dong, J., Pan, R., Zhang, W., Shen, Z., Liu, Z., Gu, D., Wang, X., Liu, X., Zhang, Q., Flint, J., Kendler, K.S., 2014. Subtypes of major depression: latent class analysis in depressed Han Chinese women. Psychol. Med. 44 (15), 3275–3288.

Lubke, G.H., Muthén, B., 2005. Investigating population heterogeneity with factor mixture models. Psychol. Methods 10 (1), 21–39.

Lubke, G.H., Miller, P.J., 2015. Does nature have joints worth carving? A discussion of taxometrics, model-based clustering and latent variable mixture modeling. Psychol. Med. 45, 705–715.

Lubke, G.H., Muthén, B., Moilanen, I.K., McGough, J.J., Loo, S.K., Swanson, J.M., Yang, M.H., Taanila, A., Hurtig, T., Järvelin, M.R., Smalley, S.L., 2007. Subtypes versus severity differences in attention-deficit/hyperactivity disorder in the Northern Finnish Birth Cohort. J. Am. Acad. Child Adolesc. Psychiatry 46 (12), 1584–1593.

Maij-de Meij, A.M., Kelderman, H., Van der Flier, H., 2005. Latent-trait latent class analysis of self-disclosure in the work environment. Multivar. Behav. Res. 40, 435–459.

Maij-de Meij, A.M., Kelderman, H., van der Flier, H., 2008. Fitting a mixture item response theory to personality questionnaire data: characterizing latent classes and investigating probabilities for improving prediction. Appl. Psychol. Meas. 32, 611–631.

Maij-de Meij, A.M., Kelderman, H., van der Flier, H., 2010. Improvement in detection of differential item functioning using a mixture item response theory model. Multivar. Behav. Res. 45, 975–999.

McKnight, P.E., Kashdan, T.B., 2009. The importance of functional impairment to mental health outcomes: a case for reassessing our goals in depression treatment research. Clin. Psychol. Rev. 29 (3), 243–259.

Mineka, S., Watson, D., Clark, L.A., 1998. Comorbidity of anxiety and unipolar mood disorders. Annu. Rev. Psychol. 49, 377–412.

Mislevy, R.J., Verhelst, N., 1990. Modeling item responses when different subjects employ different solution strategies. Psychometrika 55 (2), 195–215.
Monden, R., Stegeman, A., Conradi, H.J., de Jonge, P., Wardenaar, K.J., 2016. Predicting long-term depression outcome using a three-mode principal component model for depression heterogeneity. J. Affect. Disord. 189, 1–9.
Muthén, B., Asparouhov, T., 2006. Item response mixture modeling: application to tobacco dependence criteria. Addict. Behav. 31 (6), 1050–1066.
Muthén, L., Muthén, B., 1998–2015. Mplus User's Guide, 7th ed., Muthén & Muthén, Los Angeles.
Parker, G., Wilhelm, K., Mitchell, P., Roy, K., Hadzi-Pavlovic, D., 1999. Subtyping depression: testing algorithms and identification of a tiered model. J. Nerv. Ment. Dis. 187 (10), 610–617.
Pattyn, T., Van Den Eede, F., Lamers, F., Veltman, D., Sabbe, B.G., Penninx, B.W., 2015. Identifying panic disorder subtypes using factor mixture modeling. Depression Anxiety 32 (7), 509–517.
Pearlin, L.I., Schooler, C., 1978. The structure of coping. J. Health Soc. Behav. 19, 2–21.
Picardi, A., Viroli, C., Tarsitani, L., Miglio, R., de Girolamo, G., Dell'Acqua, G., Biondi, M., 2012. Heterogeneity and symptom structure of schizophrenia. Psychiatry Res. 198 (3), 386–394.
Rosenberg, M., 1965. The Measurement of Self-Esteem. Princeton University Press, Princeton, NJ.
Rost, J., 1990. Rasch model in latent classes: an integration of two approaches to item analysis. Appl. Psychol. Meas. 14, 271–282.
Rost, J., 1991. A logistic mixture distribution model for polychotomous item responses. Br. J. Math. Stat. Psychol. 44, 75–92.
Shafer, A.B., 2006. Meta-analysis of the factor structures of four depression questionnaires: beck, CES-D, Hamilton, and ZungJ. Clin. Psychol 62, 123–146.
Stewart, A.L., Hays, R.D., Ware, J.E., 1988. The MOS Short-Form General Health Survey: reliability and validity in a patient population. Med. Care 26, 724.
Sullivan, P.F., Kendler, K.S., 1998. Typology of common psychiatric syndromes. An empirical study. Br. J. Psychiatry 173, 312–319.
Sullivan, P.F., Kessler, R.C., Kendler, K.S., 1998. Latent class analysis of lifetime depressive symptoms in the national comorbidity survey. Am. J. Psychiatry 155 (10), 1398–1406.
Sullivan, P.F., Prescott, C.A., Kendler, K.S., 2002. The subtypes of major depression in a twin registry. J. Affect. Disord. 68 (2–3), 273–284.
Sunderland, M., Carragher, N., Wong, N., Andrews, G., 2013. Factor mixture analysis of DSM-IV symptoms of major depression in a treatment seeking clinical population. Compr. Psychiatry 54 (5), 474–483.
Tay, L., Newman, D.A., Vermunt, J.K., 2011. Using mixed-measurement item response theory with covariates (MM-IRT-C) to ascertain observed and unobserved measurement equivalence. Organ. Res. Methods 14 (1), 147–176.
Ten Have, M., Lamers, F., Wardenaar, K., Beekman, A., de Jonge, P., van Dorsselaer, S., Tuithof, M., Kleinjan, M., de Graaf, R., 2016. The identification of symptom-based subtypes of depression and an investigation of their clinical correlates and prospective course in a nationally representative cohort study. J. Affect. Disord. 190, 395–406.
Ulbricht, C.M., Rothschild, A.J., Lapane, K.L., 2015. The association between latent depression subtypes and remission after treatment with citalopram: a latent class analysis with distal outcome. J. Affect. Disord. 188, 270–277.
van Loo, H.M., Wanders, R.B.K., Wardenaar, K.J., Fried, E.I., 2017. Problems with latent class analysis to detect data-driven subtypes of depression. Mol. Psychiatry. http://dx.doi.org/10.1038/mp.2016.202.
Vermunt, J.K., Magidson, J., 2016. Technical Guide for Latent GOLD5.0: Basic, Advanced, and Syntax. Statistical Innovations Inc, Belmont.
Wanders, R.B.K., van Loo, H., Vermunt, J.K., Meijer, R.R., Hartman, C.A., Schoevers, R.A., Wardenaar, K.J., de Jonge, P., 2016. Casting wider nets for anxiety and depression: disability-driven cross-diagnostic subtypes in a large cohort. Psychol. Med. 46, 3371–3382.
Wardenaar, K.J., Van Veen, T., Giltay, E.J., Den Hollander-Gijsman, M.E., Penninx, B.W.J.H., Zitman, F.G., 2010. The structure and dimensionality of the Inventory of Depressive Symptomatology Self Report (IDS-SR) in patients with depressive disorder and healthy controls. J. Affect. Disord. 125, 146–154.
Wardenaar, K.J., Giltay, E.J., van Veen, T., Zitman, F.G., Penninx, B.W.J.H., 2012. Dimensions of the inventory of depressive symptomatology and the course and outcome of depressive and anxiety disorders. J. Psychiatr. Res. 46, 1655–1661.
Wardenaar, K.J., Monden, R., Conradi, H.J., de Jonge, P., 2015. Symptom-specific course trajectories and their determinants in primary care patients with major depressive disorder: evidence for two etiologically distinct prototypes. J. Affect. Disord. 179, 38–46.
Widiger, T.A., Clark, L.A., 2000. Toward DSM-V and the classification of psychopathology. Psychol. Bull. 126, 946–949.