Author's response to reviews

Title: Decomposing the Heterogeneity of Depression at the Person-, Symptom-, and Time-level: Latent Variable Models versus Multimode Principal Component Analysis

Authors:

Stijn de Vos (s.de.vos@umcg.nl)
Klaas J Wardenaar (k.j.wardenaar@umcg.nl)
Elisabeth H Bos (elske.bos@umcg.nl)
Ernst C Wit (e.c.wit@rug.nl)
Peter de Jonge (peter.de.jonge@umcg.nl)

Version: 5
Date: 27 May 2015

Author's response to reviews: see over
Decomposing the Heterogeneity of Depression at the Person-, Symptom-, and Time-level: Latent Variable Models versus Multimode Principal Component Analysis

Stijn de Vos MSc, Klaas J. Wardenaar PhD, Elisabeth H. Bos PhD, Ernst C. Wit PhD, Peter de Jonge PhD

1 University of Groningen, University Medical Center Groningen, Interdisciplinary Center Psychopathology and Emotion regulation, Groningen, the Netherlands

2 University of Groningen, Johann Bernoulli Institute of Mathematics and Computer Science, Groningen, the Netherlands.

§ Corresponding author: Stijn de Vos MSc, University of Groningen, University Medical Center Groningen, Interdisciplinary Center Psychopathology and Emotion regulation (ICPE), (internal mail CC-72), P.O. Box 30.001, 9700 RB, Groningen, The Netherlands, Tel: +31-50-3612101, Fax: +31-50-3619722, E-mail: s.de.vos@umcg.nl.

Email addresses:

SDV: s.de.vos@umcg.nl
KJW: k.j.wardenaar@umcg.nl
EHB: elske.bos@umcg.nl
ECW: e.c.wit@rug.nl
PDJ: peter.de.jonge@umcg.nl
Abstract

Background

Heterogeneity of psychopathological concepts such as depression hampers progress in research and clinical practice. Latent Variable Models (LVMs) have been widely used to reduce this problem by identification of more homogeneous factors or subgroups. However, heterogeneity exists at multiple levels (persons, symptoms, time) and LVMs cannot capture all these levels and their interactions simultaneously, which leads to incomplete models. Our objective is to briefly review the most widely used LVMs in depression research, illustrating their use and incompatibility in real data, and to consider an alternative, statistical approach, namely multimode principal component analysis (MPCA).

Methods

We applied LVMs to data from 147 patients, who filled out the Quick Inventory of Depressive Symptomatology (QIDS) at 9 time points. Compatibility of the results and suitability of the LVMs to capture the heterogeneity of the data were evaluated. Alternatively, MPCA was used to simultaneously decompose depression on the person-, symptom- and time-level and to investigate the interactions between these levels.

Results

QIDS-data could be decomposed on the person-level (2 classes), symptom-level (2 factors) and time-level (2 trajectory-classes). However, these results could not be integrated into a single model. Instead, MPCA allowed for decomposition of the data at the person- (3 components), symptom- (2 components) and time-level (2 components) and for the investigation of these components’ interactions.
Conclusions

Traditional LVMs have limited use when trying to define an integrated model of depression heterogeneity at the person, symptom and time level. More integrative statistical techniques such as MPCA can be used to address these relatively complex data patterns and could be used in future attempts to identify empirically-based subtypes/phenotypes of depression.

Keywords

Data cube, latent variable models, 3PCA, MPCA, heterogeneity, multimodal data, depression


**Background**

Depression is a highly prevalent and burdensome disorder that is projected to become one of the largest contributors to the global burden of disease [1]. Although research has provided considerable insight into the mechanisms underlying depression, its phenomenology and (e.g. biological/psychological/environmental) origins have remained poorly understood. To improve this situation, several problems that have hampered research thus far should be overcome. A main problem for scientific research has been the lack of validity of currently used DSM/ICD-10 depression diagnoses [2,3], which are not based on empirical evidence but on clinical consensus, leading to diagnoses with arbitrary boundaries and much overlap [4]. A second problem lies in the used syndrome approach to categorize patients, which allows for a great deal of heterogeneity among patients with the same diagnosis [3,4,5,6]. This problem is the focus of the current paper because, other than a total reconceptualization of the depression construct to improve its validity, decreasing heterogeneity can be done in currently available depression data using statistical techniques. Moreover, addressing this particular problem could already improve the specificity and interpretability of research results.

Heterogeneity is present in multiple aspects of depression: (1) the symptomatology of depression is heterogeneous, (2) the group of patients with a depression diagnosis is heterogeneous, and (3) the range of possible depressive course trajectories is heterogeneous. Consequently, it is unlikely that a few general etiological pathways could ever explain all possible patterns of depression onset, course and outcome. Although several neurobiological [7] and environmental [8, 9] factors have been shown to be generally involved in depression, observed effects have been small and hard to replicate.
Data-driven statistical methods have been used to identify more homogeneous phenotypes that could enable investigation of more phenomenology-specific etiological pathways. So far, these approaches have been used to either identify symptom subdomains or patient subgroups based on common patterns of symptomatology. Statistically, this amounts to finding latent entities within sets of symptom data obtained from depressed patients [10]. A variety of latent variable models (LVMs) have been used for this and depending on the type of data and used LVM, different aspects of depression heterogeneity can be addressed. In an ideal scenario, symptom data are obtained from the same subjects at multiple time points, yielding a dataset with a three-modal structure: a person (p) mode, having \( n \) individuals, a symptom (s) mode, representing \( m \) psychometric variables and a time (t) mode, representing \( T \) measurements over time(*). This data object can be represented mathematically as \( X = (X_{ijt}) \), where \( i \) ranges from 1 to \( n \), \( j \) from 1 to \( m \) and \( t \) ranges from 1 to \( T \). A 3-mode data structure like this is often visualized as a `data cube' (Figure 1.1) [11].

When addressing depression heterogeneity, LVMs are usually performed within a single `slice' of the data cube (see Figure 1.2-1.4; adapted from [12]). In the cross-sectional slice, the data are represented by a \( n \times m \) matrix \( X = (X_{ijk}) \) where \( k \) is a fixed point in time (Figures 1.2 and 1.3). Two LVM approaches can be used to explore the latent structure of this slice: using (1) discrete or (2) continuous LV's. When using discrete LVs, latent subgroups of depression can be discerned based on their item responses (Figure 1.3). In the case of discrete observable variables a latent class model (LCM) is used. If observable variables are continuous, a latent profile model (LPM) is used [21]. A LCM/LPM typically includes one LV and assumes it to have a finite number of levels (i.e. classes), which are observed through

\[...\]
a number of discrete variables (i.e. symptoms). To identify the optimal number of
classes, various general (e.g. Bayesian Information Criterion [BIC], Akaike
Information Criterion [AIC]) and specific [22] methods can be used. In depression
research LCM and LPM approaches have been used to identify different subgroups of
depressive patients (e.g. [10, 14,15,16,17,18,19,23]). For example, LCM identified
classes of patients with ‘severe typical’ (decreased appetite/weight), ‘severe atypical’
(increased appetite/weight) and ‘moderate severe’ depression [17,18]. In addition,
LPM identified classes of patients with different profiles on the four depressive
subdomains [23].

When using continuous LV’s, clustering of psychometric variables on
underlying factors (Figure 1.2) can be investigated. [19,20]. Factor analysis is a well-
known LVM using continuous latent variables (factors), which are measured through
observable variables (e.g. symptoms, questionnaire items). With both exploratory
factor analysis [24] and confirmatory factor analysis [25] we are interested in
explaining the variability in the observable variables (e.g. symptoms, questionnaire
data) by modelling common latent factors (e.g. depression severity). Factor analyses
have been used widely to identify homogeneous subdomains of depressive
symptomatology [19] and/or to investigate measurement scales’ structure/construct
validity [20,26].

In the longitudinal slice of the data cube (Figure 1.4), some variable $p_j$ is fixed
and responses from all individuals over all time points are considered. The data is
represented by a $n \times T$ matrix $X = (X_{ijk})$ for fixed $j$. Two aspects are of interest
when accounting for heterogeneity in this data: (1) the modelling of individual growth
curves and (2) evidence for the presence of subpopulations with respect to their
growth. These models can be used to model the heterogeneity in the course
trajectories of depression (e.g. [27]). Previous studies have mostly used discrete LVs
(latent class growth models [LCGMs] and growth mixture models [GMMs]) and
continuous outcomes to identify distinct depressive course groups (e.g. ‘quick
remission’ vs. ‘chronic course’ [27,28,29,30,31]). With LCGMs, each person’s
trajectory on a continuous outcome (e.g. depression severity) is assumed to belong to
one of a number of classes. Each class has its own growth-curve characteristics (e.g.
intercept and slope means). A GMM is similar to a LCGM but also models within-
class heterogeneity by estimating growth-parameter (co)variances [32]. This added
flexibility often leads GMMs to be more parsimonious and to fit better to the data
than LCGMs [27, 32].

It is important to note that the results of above described techniques in
different data slices yield incompatible results in terms of the modes in which
heterogeneity is explained and in which homogeneity is assumed [12]. Consequently,
the results of traditional LVM studies can be very informative about the heterogeneity
in particular modes of the data, but cannot be directly integrated into a single
framework to explain all depression heterogeneity: i.e. factor analytic results about
symptom-domains do not inform us about the best way to subgroup patients and vice
versa. Because traditional LVMs fail to account for simultaneous variability in
multiple modes of depression data, current models of depression heterogeneity are
still incomplete and/or oversimplified.

A possible solution to this problem is to use statistical techniques that allow
for simultaneous decomposition of depression on the person-, symptom- and time-
level. This can be done with multiway principal component analysis (MPCA) [33].
MPCA is a higher-dimensional generalization of regular PCA that can be used to
detect clustering tendencies in higher-dimensional datasets. Because we have three
modes, we are interested in PCA models that work in three dimensions. There are various models that perform PCA in three dimensions, one of which is the Tucker3 model. This model is defined mathematically as

\[
X_{ijk} = \sum_{p=1}^{P} \sum_{q=1}^{Q} \sum_{r=1}^{R} a_{ip} b_{jq} c_{kr} g_{pqr} + e_{ijk}
\]

where \(X_{ijk}\) is the entry in the three dimensional data set \(X\) at place \((i,j,k)\) and \(a_{ip}\), \(b_{jq}\), \(c_{kr}\) are the elements of the component matrices \(A\), \(B\) and \(C\), respectively. The numbers \(g_{pqr}\) belong to the core array \(G\) and \(e_{ijk}\) represents the error of data entry \(x_{ijk}\). If the data set \(X\) contains data from \(n\) subjects, \(m\) symptoms and \(T\) time points and \(P\), \(Q\), \(R\) are the number of components in the A-, B- and C-mode respectively, then \(A\) has size \(n \times P\), \(B\) has size \(m \times Q\) and \(C\) has size \(T \times R\). The core array is of size \(P \times Q \times R\). See figure 2 for a graphical representation of the Tucker3 model.

The component matrices \(A\), \(B\) and \(C\) represent the loadings of the data entries onto their respective components. This can be compared to standard PCA, where, after picking a number of components, component scores are estimated. One difference between threeway PCA models and standard PCA is the core array \(G\). This array contains numbers that represent how the components in the A-, B- and C-modes interact. It is this array that enables the integrated modeling of heterogeneity in three dimensional data. Without it, there would be no assumption of threeway interactions.

The interpretation of 3PCA component scores is similar as in the case of standard PCA. The symptom-mode components represent a characterization of groups of symptoms, the time-mode components illustrate groups of time trajectories and the person-mode components represent characterizations of groups of persons. The latter means that every person in the 3-mode dataset has a profile of scores on the person-mode components. In the context of depression research, these scores characterize the...
way in which in this person the symptom-components behave over time. Thus, the person-mode component scores provide person-specific characterizations, which can be interpreted in a similar way as, for instance, personality assessments where a person’s scoring profile on multiple traits gives a quantitative characterization of his personality.

Other three-way PCA models are obtained when constraints are imposed on the number of components per mode and/or the shape of the core array, yielding models such as the CANDECOMP/PARAFAC model. For an expansive overview of various multiway PCA models (including Tucker3), see [33].

This paper aims to illustrate and discuss the abovementioned problems and present the use of 3PCA as a possible solution. First, each of the LVMs is performed in an example depression dataset to illustrate their typical results. Second, the more integrative approach of 3PCA is demonstrated and its results are compared with the results of the traditional LVMs.
Methods

To compare the results from all the above mentioned LVM techniques, they were run in a real-life, longitudinal depression dataset. EFA and LCA were conducted on the cross-sectional slice of this dataset and LCGMs and GMMs were applied to the longitudinal slice. 3PCA was performed on the whole dataset.

Participants

The data came from 147 outpatients visiting a specialized day-care depression unit at the Department of Psychiatry at the University Medical Center Groningen (UMCG). The data collection was part of a 16 weeks long treatment program. The mean age was 42.9 (SD=11.9), 50.8% were female, and none were hospitalized. Patients were administered the Quick Inventory of Depressive Symptomatology (QIDS-SR; [34]) at baseline and, when possible, during every following week. The 147 patients with complete baseline data formed the cross-sectional data-slice (Figure 1.2 and 1.3) and were used for EFA and LCM. A subsample of 82 patients (mean age=42.3 [sd = 12.1]; 51.2% female) with at least nine complete weekly follow-up measurements was used for all longitudinal analyses (LCGM, GMM and 3PCA). For the EFA and LCM, baseline cross-sectional item-level QIDS-SR data were used. For the LCGM and GMM, aggregate QIDS-SR sum-scores at repeated time points were used. For 3PCA item-level QIDS data at each of the nine repeated time points were used.

The demographic data of the sample is summarized in table 1.

Participant approval

The data collection was part of a treatment program for patients at the Department of Psychiatry at the University Medical Center Groningen. At intake, patients are informed that the data collection is part of the general policy of the UMCG to monitor
treatment outcome, that outcomes are made available only to their therapist and that
the data will be used for research purposes, but only in anonymized form. If patients
object to such use, their data are removed. The Dutch Central Committee on Research
Involving Human Subjects (CCMO) approved the regulations and agreed with this
policy. The study was conducted in accordance with the Declaration of Helsinki.

**Statistical guidelines**

Reporting guidelines have been followed where applicable.

**Measurements**

The QIDS consists of 16 items, coded on a 4-point Likert scale (0,1,2,3). Because the
highest category (3) was endorsed very scarcely, this category was merged with
category 2, resulting in a 3-point scale (0,1,2), which was more suitable for
conducting LVMs with categorical data in Mplus.

Prior to analyses, hyposomnia items 1-3 were recoded into one item (item 1-3)
using the highest score. Mutually exclusive items on increased and decreased
appetite/weight were recoded into single appetite-change (item 6/7) and weight-
change (item 8/9). For the longitudinal analyses, QIDS sum scores were computed for
the nine consecutive time-points by adding up the item-scores at each time-point after
recoding (item 1-3, item 6/7 and item 8/9).

**Statistical analyses**

**Latent Variable Models**

All analyses (EFA, LCM, LCGM, GMM) were conducted with Mplus, version 6.12.
[35]. Model estimation was done with Maximum Likelihood with robust standard
tests (MLR). For each analysis, differently specified models (e.g. different class-
numbers) were run. Model selection was performed by comparing the AIC and BIC, two common model selection criteria, between competing models.

EFA was conducted using the baseline QIDS-SR item-scores. During EFA, the optimal number of factors was determined by comparing the AIC and BIC between models with different numbers of factors. The default rotation method in Mplus (oblique Geomin) was used and the rotated factor loadings were inspected to interpret the content/meaning of each of the factors. Items were allocated to the factor on which they showed the highest factor loading. There exist other model selection methods such as the Scree-plot and parallel analysis. However, as the number of extracted factors was not the main point of interest in this study, these additional techniques were not considered. For the fitting of the LCMs, which was done using the baseline QIDS-SR item-scores, the profiles of item-endorsement probabilities in each class were plotted to interpret class differences.

LCGM and GMM fitting was done using the aggregate QIDS-SR sum scores at the 9 repeated time points. In both analyses, class-specific intercepts, slopes and quadratic terms were estimated. For LCGM, the variances of these terms were fixed at zero. For the GMMs random intercepts were estimated and slopes/quadratic terms were fixed.

### 3PCA

A 3PCA was performed in the longitudinal dataset (n=82) using the item-level QIDS-data at each of the nine follow-up waves. The analyses were conducted with R package “ThreeWay” [36]. Data pre-processing (centering and normalization) and fit-percentage calculation was done following [37]. The pre-processing described there allows us to use the 3PCA analysis to investigate heterogeneity up-and-above the
general mean trend in the dataset, which is shared by all patients ([36, p. 6]). The
ThreeWay R-package presents the user with modeling choices that have to be made
during the analysis process, such as the number of components to be estimated per
mode. There is also an option to calculate fit percentage for a variety of component
combinations, after which the preferred number of components per mode can be
selected. As is the case in standard PCA, model selection is non-trivial because a
higher fit percentage can always be achieved by selecting more components. Taking
into account fit percentage and the number of parameters in the model, we selected a
model with, respectively 3, 2 and 2 components for the person-, symptom- and time-
mode. The 3PCA model (the number of components per mode) was selected based on
the percentage of explained variance.

The used R and Mplus scripts are available in additional file 1 (Mplus scripts)
and additional file 2 (R scripts) to this paper.
**Results**

**EFA**

EFA results (Table 2) showed that based on the BIC, a 1-factor model would be deemed optimal; all QIDS items loaded high on one factor (Table 3), except for hypersomnia (item 4) and weight change (item 8/9) suggesting potential multidimensionality. Indeed, a decreasing AIC suggested that a second factor could be added. In the 2-factor model, the hypersomnia (item 4) and weight change items loaded on one ‘vegetative’ factor and all other items loaded on a second ‘mood/cognitive’ factor. The AIC favoured addition of a third factor, but this model had two factors with only one item loading on it. These results showed that symptomatology could be described by a 1-factor model, but that a 2-factor model was more informative in terms of differentiation between mood/cognitive and vegetative domains.

**LCM**

LCM results are shown in Table 2. The 2-class model has the lowest BIC, whereas the AIC continued to decrease with each class addition. The 2-class model (Figure 3) indicated that heterogeneity among persons could be explained by the existence of a low and a high severity class. Motivated by the AIC values, a 3-class model was also investigated, which showed a low-severity class (16.8%), but further distinguished two more qualitatively different high-severity classes (see Figure 3): one class (58.7%) showed higher probabilities of endorsing suicidality, loss of interest, and psychomotor problems than the other high-severity class (24.5%). The 4-class model showed additional qualitative differentiation between high-severity classes, with one
class showing a relatively higher probability of suicidal ideation (51.2%) and another
of relatively higher psychomotor agitation (15.0%; see Figure 3).

**LCGM and GMM**

GMM models fit better to the data than LCGM models as shown by lower AIC and
BIC levels. In the GMM the AIC and BIC decreased until addition of a fourth class.
However, the best fitting 3-class model had one small class (n=3) that could not be
reliably interpreted. Therefore, the 2-class model was selected for further inspection.
The observed trajectories are shown in Figure 4: one class (88.2%) was characterized
by a relatively quick decrease in symptomatology compared to the other class, which
showed a slower decrease in severity.

**3PCA**

The results of the 3PCA analyses are shown in tables 4, 5 and 6. Because it showed
comparatively good fit to the data (explained variance=80.3%), a 3PCA model with
three components in the person-mode, two in the symptom-mode and two in the time-
mode was selected (3,2,2). The scores on the symptom components (Table 4) showed
that the symptom mode could be decomposed into a somatic/affective and
cognitive/appetitive component. The scores on the time components (Table 5)
indicated that the initial 5 time points scored high on one component and time points
6 through 9 scored high on the a second time component. Considering that these
scores represent variability up and above the general downward trend in the data, the
first component was interpreted to represent a ‘persisting’ time component while the
second component represented an ‘improving’ time component.

The characteristics of the person-mode components could be deduced from the
core-array (Table 6), which quantifies the interactions between the symptom- and
time-mode components per person-mode. The first person-mode was characterized by comparatively high somatic/affective and cognitive/appetitive symptom-component scores in the first, ‘improving’ time-phase and much lower symptom-component scores in the ‘persisting’ time-phase. As such, persons with high scores on the first person-mode component were characterised by ‘overall quick recovery’. The second person-mode component was characterized by much higher somatic affective than ‘cognitive/appetitive’ symptom-component scores in the ‘improving’ time-phase and this situation remained the same for the ‘persisting’ time-phase. Therefore, scores on this person-mode components were associated with ‘persistent somatic/affective symptomatology’. The third person-mode component was characterized by slightly different ‘somatic/affective’ and ‘cognitive/appetitive’ scores in the ‘improving’ time-phase, which both increased in the ‘persisting’ time-phase. Scores on this component were thus associated with ‘overall increasing symptomatology’.

LVMs versus 3PCA

Some of the traditional LVM results showed parallels with some aspects of the 3PCA results. The 3-PCA symptom-mode components showed quite some overlap with the results of the EFA: items 1-3, 5, 10 and 13-16 were grouped similarly in both models, although the component scores of items 11 and 12 (appetite change and weight change) in the 3PCA model, made the interpretation different from that of the EFA model.

The time-mode components cannot be directly compared to the GMA because the former is interpreted as two components of the mode time, whereas the latter is interpreted in terms of depression growth-curves over time. In as similar vein, the person-mode components cannot be compared to the LCM results because the former
characterize persons in terms of symptom-component scores over time, whereas the
latter characterizes persons based only on their symptom-score profiles. As such, the
core-array of the 3PCA adds an integrative aspect to the analyses that is absent from
the traditional techniques and allows for the description of person-heterogeneity in
terms of symptom-scores across time.
Discussion

Although LVMs can be used to address the heterogeneity of depression in particular slices of a three-modal data set, LVMs used so far in depression research are unsuited to form a complete picture of depression heterogeneity, integrating all sources of heterogeneity at the same time. As such, traditional LVMs are suboptimal to find a solution to the heterogeneity problem. Aside from the incompatibility of the results obtained from different LVMs and different data-slices, the assumptions of local independence of the symptoms and the independence of the different data-slices are unrealistic.

An alternative for traditional LVMs was presented in a 3PCA approach, which describes heterogeneity in three-modal data in a more integrated manner by simultaneously decomposing each mode into several components and considering the interactions between the different sources of heterogeneity. Although decomposing a single mode into a number of components by itself is not an improvement over more traditional LVMs, 3PCA has the added value that it allows for a quantification of the relationship between the various modes’ components. Since heterogeneity is difficult to describe by looking at any mode on its own, the core array offers interesting new insight into the heterogeneity of depression. When applied in research, this means that, rather than just showing how depressive symptomatology can be decomposed into symptom domains or showing how a sample can be decomposed into subtypes based on symptom profiles, 3PCA provides a more integrated insight into how persons differ in terms of their course-trajectories on different symptom-domains.

Although 3PCA results can sometimes be hard to interpret, the presented results clearly illustrate its use to gain more insight into the nature and extent of heterogeneity among depressive patients in a given sample, providing an integrated
picture of how patients differ from each other in terms of their course-trajectories on different symptom domains.

Limitations

The results of the current study should be interpreted taking into account some study limitations. Firstly, the sample size (N = 82) for the longitudinal analyses is arguably small, while the number of time points is limited (T = 9). This is a consequence of the number of participants having nine or more measurements; only 82 of them had nine measurements. Consequently, model results might be influenced by the difference in sample sizes. Second, there exist numerous ways to perform traditional LV analyses. For example, it is conceivable that a different rotation method would have resulted in somewhat different factor loadings, which would also affect the comparability with the 3PCA symptom-mode components. Third, the factor analysis and latent class analysis was performed on data coming from the 147 patients, while the longitudinal analyses were performed on data from a subset of those patients (n = 82). The latter was due to the fact that in its current form, 3PCA cannot handle missing data.

Future studies should investigate 3PCA in larger samples and incorporate a larger number of symptoms and follow-up measurements. Moreover, the scientific and/or clinical relevance of 3PCA components should be evaluated.

Conclusions

Investigation of heterogeneity by using LVMs cannot provide insight into all sources of heterogeneity simultaneously. The complexity of the problem requires the development of techniques that investigate patient subdivision based on their
symptomatology and course. To this end, MPCA might prove to be a valuable alternative.

**Availability of supporting data**

The data used in this study is available from the corresponding author upon request.

**List of abbreviations used**

DSM = Diagnostic and Statistical Manual; ICD-10=International Classification of Disease 10; LVM = latent variable model; LCM = latent class model; LPM=Latent Profile Model; LCGM = latent class growth model; GMM = growth mixture model; EFA = exploratory factor analysis; CFA=Confirmatory Factor Analysis; MPCA = multiway principal component analysis; 3PCA=three-way principal component analysis.

**Competing interests**

None of the authors report any conflict of interest.

**Authors' contributions**

SDV drafted the manuscript. KJW, EHB, PDJ and ECW made substantial contributions to the manuscript, study concept and design.

**Acknowledgements**

We thank R. Monden for providing statistical advice and R.W. Heins for providing the research data.
The work by SDV, KW, EB and PDJ was funded by a VICI grant (no. 91812607) awarded to PDJ by the Netherlands Organization for Scientific Research (NWO-ZonMW). ECW received no external funding for his work on the reported study.

References


33 Kroonenberg PM. Applied multiway data analysis. Hoboken NJ: John Wiley & Sons; 2008


536

537
Figures

Figure 1. Slices of the data cube. Illustration of the data cube, its different slices and associated techniques. p=person, s=symptom, t=time point.

Figure 2. The elements of a 3-mode principle component model. Analysis of a 3-mode dataset yields a model consisting of a person-mode component matrix (A), a symptom-mode component matrix (B) and a time-mode component matrix (C). In addition, the interactions between the different modes’ components are described by the core-array (G).

Figure 3. LCA item probabilities. Item probabilities for 2-, 3-, and 4-class latent class models in a sample of 147 help-seeking patients. The y-axis denotes the probability of endorsing a non-zero response.

Figure 4. GMM trajectories. Class-specific trajectories of observed mean Quick Inventory of Depressive Symptoms (QIDS) scores for the best-fitting (GMM) in a sample of 82 help-seeking patients.
### Table 1: sample descriptives

<table>
<thead>
<tr>
<th>Data</th>
<th>Sample size</th>
<th>Mean age (s.d.)</th>
<th>% female</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cross sectional</td>
<td>147</td>
<td>42.9 (11.9)</td>
<td>50.3</td>
</tr>
<tr>
<td>Longitudinal</td>
<td>82 (9 time points)</td>
<td>42.3 (12.1)</td>
<td>51.2</td>
</tr>
</tbody>
</table>

s.d.=standard deviation
Table 2: Model assessment with Maximum Likelihood with robust standard errors (MLR)

<table>
<thead>
<tr>
<th>Analysis</th>
<th>Degrees of freedom</th>
<th>AIC</th>
<th>BIC</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>EFA</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(n=147)</td>
<td>1-factor</td>
<td>36</td>
<td>3093</td>
</tr>
<tr>
<td></td>
<td>2-factor</td>
<td>47</td>
<td>3090</td>
</tr>
<tr>
<td></td>
<td>3-factor</td>
<td>57</td>
<td>3082</td>
</tr>
<tr>
<td><strong>LCM</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(n=147)</td>
<td>1-class</td>
<td>24</td>
<td>3460</td>
</tr>
<tr>
<td></td>
<td>2-class</td>
<td>49</td>
<td>3161</td>
</tr>
<tr>
<td></td>
<td>3-class</td>
<td>74</td>
<td>3097</td>
</tr>
<tr>
<td></td>
<td>4-class</td>
<td>99</td>
<td>3071</td>
</tr>
<tr>
<td></td>
<td>5-class*</td>
<td>124</td>
<td>3052</td>
</tr>
<tr>
<td><strong>LCGM</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(n=82)</td>
<td>1-class</td>
<td>12</td>
<td>4670</td>
</tr>
<tr>
<td></td>
<td>2-class</td>
<td>16</td>
<td>4363</td>
</tr>
<tr>
<td></td>
<td>3-class</td>
<td>20</td>
<td>4267</td>
</tr>
<tr>
<td></td>
<td>4-class</td>
<td>24</td>
<td>4220</td>
</tr>
<tr>
<td></td>
<td>5-class*</td>
<td>28</td>
<td>4173</td>
</tr>
<tr>
<td><strong>GMM</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(n=82)</td>
<td>1-class</td>
<td>15</td>
<td>4188</td>
</tr>
<tr>
<td></td>
<td>2-class</td>
<td>21</td>
<td>4163</td>
</tr>
<tr>
<td></td>
<td>3-class</td>
<td>27**</td>
<td>4146</td>
</tr>
<tr>
<td></td>
<td>4-class</td>
<td>33**</td>
<td>4158</td>
</tr>
</tbody>
</table>

*) smallest class contains < 10 subjects

**) smallest class contains < 5 subjects

557
558
Table 3: Rotated factor loadings of the 1- and 2-factor model

<table>
<thead>
<tr>
<th>QIDS item number</th>
<th>Item label</th>
<th>1-factor model</th>
<th>2-factor model</th>
<th>‘Mood/Cognitive’</th>
<th>‘Vegetative’</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-3</td>
<td>Hyposomnia</td>
<td>0.35</td>
<td>0.37</td>
<td>-0.23</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Hypersomnia</td>
<td>0.14</td>
<td>0.00</td>
<td>0.85</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Feeling sad</td>
<td>0.89</td>
<td>0.88</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>6/7</td>
<td>Changed appetite</td>
<td>0.40</td>
<td>0.36</td>
<td>0.25</td>
<td></td>
</tr>
<tr>
<td>8/9</td>
<td>Changed weight</td>
<td>0.12</td>
<td>0.06</td>
<td>0.36</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Concentration</td>
<td>0.85</td>
<td>0.88</td>
<td>-0.13</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>View of myself</td>
<td>0.69</td>
<td>0.67</td>
<td>0.06</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>Thoughts of death/suicide</td>
<td>0.66</td>
<td>0.61</td>
<td>0.32</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>General interest</td>
<td>0.78</td>
<td>0.79</td>
<td>-0.07</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>Energy level</td>
<td>0.69</td>
<td>0.69</td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>Psychomotor retardation</td>
<td>0.56</td>
<td>0.56</td>
<td>0.02</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>Psychomotor agitation</td>
<td>0.39</td>
<td>0.41</td>
<td>-0.08</td>
<td></td>
</tr>
</tbody>
</table>

QIDS=Quick Inventory of Depressive Symptomatology
Loadings after oblique Geomin rotation.
Table 4: 3PCA symptom component scores

<table>
<thead>
<tr>
<th>QIDS item number</th>
<th>Item</th>
<th>1: somatic/affective</th>
<th>2: cognitive/appetitive</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-3</td>
<td>Hyposomnia</td>
<td>0.36</td>
<td>-0.19</td>
</tr>
<tr>
<td>4</td>
<td>Hypersomnia</td>
<td>-0.15</td>
<td>0.21</td>
</tr>
<tr>
<td>5</td>
<td>Feeling sad</td>
<td>0.38</td>
<td>0.12</td>
</tr>
<tr>
<td>6/7</td>
<td>Changed appetite</td>
<td>-0.01</td>
<td>0.63</td>
</tr>
<tr>
<td>8/9</td>
<td>Changed weight</td>
<td>0.03</td>
<td>0.42</td>
</tr>
<tr>
<td>10</td>
<td>Concentration</td>
<td>0.39</td>
<td>0.00</td>
</tr>
<tr>
<td>11</td>
<td>View of myself</td>
<td>0.16</td>
<td>0.28</td>
</tr>
<tr>
<td>12</td>
<td>Thoughts of death/suicide</td>
<td>0.13</td>
<td>0.36</td>
</tr>
<tr>
<td>13</td>
<td>General interest</td>
<td><strong>0.43</strong></td>
<td>-0.07</td>
</tr>
<tr>
<td>14</td>
<td>Energy level</td>
<td><strong>0.30</strong></td>
<td>0.14</td>
</tr>
<tr>
<td>15</td>
<td>Psychomotor retardation</td>
<td>0.25</td>
<td>0.19</td>
</tr>
<tr>
<td>16</td>
<td>Psychomotor agitation</td>
<td><strong>0.41</strong></td>
<td>-0.25</td>
</tr>
</tbody>
</table>

3PCA=three-mode principle component analysis; QIDS=Quick Inventory of Depressive Symptomatology
### Table 5: 3PCA time component scores

<table>
<thead>
<tr>
<th>Time points</th>
<th>1: persisting phase</th>
<th>2: improving phase</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-0.19</td>
<td>0.48</td>
</tr>
<tr>
<td>2</td>
<td>-0.05</td>
<td>0.50</td>
</tr>
<tr>
<td>3</td>
<td>0.08</td>
<td>0.45</td>
</tr>
<tr>
<td>4</td>
<td>0.11</td>
<td>0.40</td>
</tr>
<tr>
<td>5</td>
<td>0.18</td>
<td>0.35</td>
</tr>
<tr>
<td>6</td>
<td><strong>0.35</strong></td>
<td>0.16</td>
</tr>
<tr>
<td>7</td>
<td><strong>0.49</strong></td>
<td>-0.02</td>
</tr>
<tr>
<td>8</td>
<td><strong>0.55</strong></td>
<td>-0.10</td>
</tr>
<tr>
<td>9</td>
<td><strong>0.50</strong></td>
<td>-0.04</td>
</tr>
</tbody>
</table>

3PCA=three-mode principle component analysis
### Table 6: 3PCA core component scores

<table>
<thead>
<tr>
<th>Person components</th>
<th>Symptom component</th>
<th>Time component</th>
<th>1: somatic/affective</th>
<th>2: cognitive/appetitive</th>
</tr>
</thead>
<tbody>
<tr>
<td>1: early recovery</td>
<td>1: improving phase</td>
<td>24.55</td>
<td>10.38</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2: persisting phase</td>
<td>5.80</td>
<td>-0.89</td>
<td></td>
</tr>
<tr>
<td>2: persistent somatic/affective</td>
<td>1: improving phase</td>
<td>6.05</td>
<td>-18.89</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2: persisting phase</td>
<td>2.75</td>
<td>-15.89</td>
<td></td>
</tr>
<tr>
<td>3: increasing symptoms</td>
<td>1: improving phase</td>
<td>16.61</td>
<td>4.73</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2: persisting phase</td>
<td>28.60</td>
<td>10.50</td>
<td></td>
</tr>
</tbody>
</table>

3PCA = three-mode principle component analysis
Endnotes

(*): what we here refer to as a mode is known as a dimension in mathematics. This is in contrast with medical jargon, where a mode (e.g. symptom mode) contains various dimensions (e.g. positive affect, negative affect).

Additional Files

Additional file 1: a pdf file containing the Mplus scripts for the presented latent variable models.

Additional file 2: a pdf file containing the R script for the presented multiway principle component analysis.