ORIGINAL ARTICLE

Revealing the dynamic network structure of the Beck Depression Inventory-II

L. F. Bringmann[1]*, L. H. J. M. Lemmens[2], M. J. H. Huibers[2,3], D. Borsboom[4] and F. Tuerlinckx[1]

[1] *Faculty of Psychology and Educational Sciences, University of Leuven, Leuven, Belgium*

[2] *Department of Clinical Psychological Science, Maastricht University, Maastricht, The Netherlands*

[3] *Department of Clinical Psychology, VU University of Amsterdam, Amsterdam, The Netherlands*

[4] *Department of Psychology, University of Amsterdam, Amsterdam, The Netherlands*

Background. Structured interviews and questionnaires are important tools to screen for major depressive disorder. Recent research suggests that, in addition to studying the mean level of total scores, researchers should focus on the dynamic relations among depressive symptoms as they unfold over time. Using network analysis, this paper is the first to investigate these patterns of short-term (i.e. session to session) dynamics for a widely used psychological questionnaire for depression – the Beck Depression Inventory (BDI-II).

Method. With the newly developed vector autoregressive (VAR) multilevel method we estimated the network of symptom dynamics that characterizes the BDI-II, based on repeated administrations of the questionnaire to a group of depressed individuals who participated in a treatment study of an average of 14 weekly assessments. Also the centrality of symptoms and the community structure of the network were examined.

Results. The analysis showed that all BDI-II symptoms are directly or indirectly connected through patterns of temporal influence. In addition, these influences are mutually reinforcing, 'loss of pleasure' being the most central item in the network. Community analyses indicated that the dynamic structure of the BDI-II involves two clusters, which is consistent with earlier psychometric analyses.

Conclusion. The network approach expands the range of depression research, making it possible to investigate the dynamic architecture of depression and opening up a whole new range of questions and analyses. Regarding clinical practice, network analyses may be used to indicate which symptoms should be targeted, and in this sense may help in setting up treatment strategies.

Introduction

Major depressive disorder (MDD) is a complex and burdensome mental health disorder made up of a wide variety of symptoms (APA, 2000; WHO, 2001; Kessler *et al.* 2003; Hardeveld *et al.* 2010). Structured interviews and questionnaires, such as the Hamilton Depression Rating Scale (HAMD; Hamilton, 1960) and Beck Depression Inventory (BDI; Beck *et al.* 1961; BDI-II; Beck *et al.* 1996*b*) are important and commonly used tools to screen for, study, and follow the course of MDD (Beck *et al.* 1996*a*). In longitudinal studies, a total score, which results from simply adding all symptom scores, is often used as a measure of changes in depression severity. Relatively few studies use a more fine-grained analysis, in which the reduction of

depression severity is studied by examining specific (clusters of) symptoms of depression instead of using the total score (e.g. Bhar *et al.* 2008; Stewart & Harkness, 2012; Fournier *et al.* 2013).

What all the above studies have in common is that they are based on the latent variable model. According to this model, symptoms of a given disorder are assumed to share an essential property; namely, their causal dependence on a latent variable, from which all symptoms arise (Borsboom, 2008; Kendler *et al.* 2011). In this perspective, symptoms experienced by patients are merely effects of the relevant latent variable (in this case depression). Standard models assume the symptoms to be statistically independent given the latent variable, and as a result, symptom associations are viewed to be spurious (Borsboom, 2008). Specifically, in the standard model, symptoms are not considered to have autonomous influence on one another. The latent variable approach, as utilized in standard models, is therefore not suitable to examine the dynamic relations between symptoms.

* Address for correspondence: L. F. Bringmann, Department Quantitative Psychology and Individual Differences, University of Leuven, Tiensestraat 102 – Box 3713, 3000 Leuven, Belgium.

(Email: laura.bringmann@ppw.kuleuven.be)

The recently developed network approach (Cramer et al. 2010) steps away from this latent variable model by proposing that research should no longer focus exclusively on the *mean level* of symptoms or change therein (e.g. an overall score or a reduction of symptoms). Instead, this approach emphasizes that clinical research should also focus on the relation between individual symptoms from one time point to another, which we denote here as *short-term dynamics*.†[1]

The importance of examining these short-term dynamics is supported by various sources. First, theories of treatments for depression tend to focus on the short-term symptom dynamics when describing their proposed mechanisms of change. For example, according to Beck's et al. (1979) cognitive theory, change in cognitive processes (e.g. negative thinking) leads to changes in symptoms such as affect (Beck, 1964; Rush et al. 1981). Second, in clinical practice it is commonly observed that if patients experience relief in one symptom (e.g. sleeping problems), other symptoms start to wane as well, indicating the start of recovery; this is especially notable when symptoms are systematically assessed at the start of each therapy session, as is the case in cognitive therapy (Beck et al. 1979). Third, recent studies indicate that depression risk factors and stressful life events have differential effects on depressive symptoms (Cramer et al. 2012; Fried et al. 2013). As Cramer et al. (2012) showed, correlations between symptoms were directly influenced by the stressful life events and could not be explained by changes in an underlying common cause, in this case the risk to develop depression. This further supports the idea that symptoms have an autonomous influence on one another. Being able to objectively describe such symptom-by-symptom interactions can give important clues for clinical research and practice.

Apart from their substantive plausibility, network approaches open up a new range of research questions. For example, estimating a network of symptoms from depression questionnaires allows for an objective assessment of the centrality of symptoms (Boccaletti et al. 2006; Opsahl et al. 2010). Symptoms with a central position in the network are probably the most important or influential ones and are therefore likely to cause the symptom spread to continue. Studying these central symptoms can give clues for further clinical research. One could investigate, for instance, the commonly held assumption that anhedonia (loss of pleasure and interest) and depressed mood are central symptoms of depression as stated by the most prevalent diagnostic systems DSM-IV-TR (APA, 2000) and

ICD-10 (WHO, 2008). In addition, once the network has been estimated, the community structure of the network can be examined (Girvan & Newman, 2002). A community is present if some clusters of symptoms are more strongly interconnected with each other than with symptoms that are not part of the cluster. In this way, the dynamic architecture of depression can be investigated.

This paper will be the first to investigate the short-term dynamics of one of the most widely used psychological questionnaires for depression: the BDI-II. Inspired by the possibilities of the network approach, we will apply a novel method developed by Bringmann et al. (2013) that is able to explore these symptom dynamics, and infer a network structure of BDI-II symptoms. Until recently, it was not possible to infer these kinds of directed and weighted networks from clinical questionnaires since two important requirements for studying short-term dynamics, intensive longitudinal data in which a set of symptoms is measured frequently across time, and a suitable statistical method, were lacking. Intensive longitudinal data is still sparse, but in a recent study by Lemmens et al. (2011; Lemmens et al. unpublished observations) such data for the BDI-II were collected. Second, the newly developed vector autoregressive (VAR) multilevel method, which is a combination of multilevel (hierarchical) and time-series models, is suited for analysing these kinds of clinical longitudinal data. These data have rather short time series (~20 time points) for a large sample of patients (Bringmann et al. 2013). Note that since only few studies have investigated single specific (clusters of) symptoms of the BDI-II or even interactions between symptoms in general, all analyses are exploratory.

The structure of the paper is as follows: first, we will infer the network(s) representing the short-term dynamics of BDI-II symptoms. Second, we will study the centrality of symptoms. Based on the DSM-IV and ICD-10, one would expect the BDI-II items that are intuitively most closely related to the main symptoms, anhedonia and depressed mood (namely items: 'loss of interest', 'loss of pleasure' and 'sadness'), to be the most central ones in the network(s). In the third and last part, we will analyse whether communities are present in BDI-II network(s). Since the network(s) consists of a fair number of symptoms (i.e. 21), we expect the emergence of new clusters of symptoms or community structures.

Method

Data

The data in the current study come from a large randomized clinical trial (RCT), which examined the

---

† The notes appear after the main text.

effectiveness, relapse prevention and mechanisms of change of cognitive therapy (CT) *v.* interpersonal psychotherapy (IPT) for depression (Lemmens *et al.* 2011; Lemmens *et al.* unpublished observations). In this study, 182 patients (aged between 18 and 65 years) with a DSM-IV diagnosis of MDD were randomly allocated to one of three conditions: (*a*) CT (*n*=76), (*b*) IPT (*n*=75), or (*c*) an 8-week waiting-list control (WLC) condition followed by treatment of choice (CT or IPT, *n*=31). In the current study, we did not differentiate between patients who started therapy immediately and who started after 8 weeks. This resulted in a sample size of 99 for the CT condition (mean age=40 years, S.D.=12 years, 80% female) and a sample size of 83 for the IPT condition (mean age=41 years, S.D.=12 years; 64% female). There were no significant differences in demographic and clinical characteristics between the groups. Each patient participated in 3–20 weekly individual sessions, depending on the progress of the patient or due to drop out. On average, patients completed 14 sessions (S.D.=5).[2] The BDI-II was administered before each session to assess depression severity. Of the 2661 sessions, 2.5% of the BDI-II data were missing. Further details concerning the design of the trial and effectiveness of the interventions have been fully reported elsewhere (Lemmens *et al.* 2011; Lemmens *et al.* unpublished observations).

BDI-II

The BDI-II (Beck *et al.* 1996*b*; Dutch translation by van der Does, 2002) is one of the most widely used and empirically validated questionnaires for screening depression. The BDI-II is a self-report questionnaire measuring the severity of depression with 21 items. Each item is rated on a 4-point Likert-scale ranging from 0 to 3. The total score, ranging from 0 to 63, is constructed by adding the item scores, with higher scores reflecting more severe depressive symptomatology.

Interventions

CT and IPT are two of the most empirically validated psychotherapies used for treating depression (Hollon *et al.* 2002; Cuijpers *et al.* 2008, 2011). CT is based on Beck's *et al.* (1979) cognitive theory, which states that depression results from maladaptive information-processing strategies that are maintained by dysfunctional behavioural responses. CT focuses on identifying and changing dysfunctional cognitions, schemas and attitudes in order to treat depression. In IPT, the interpersonal model of depression is central (Klerman *et al.* 1984). According to this model, major disturbances in the interpersonal domain may cause and maintain depression. It is assumed that depressive symptoms can be reduced through the improvement of interpersonal functioning.

Statistical analysis

The BDI-II network

First, we inferred the BDI-II network by analysing the short-term dynamics between the 21 symptoms across the 20 weeks of therapy with a modified version of the multilevel VAR method (Bringmann *et al.* 2013).[3] In the multilevel VAR method, the time dynamics between the 21 symptoms of the BDI-II from one moment to the other are represented by a VAR model (see also Tschacher *et al.* 2012 for a similar approach). In the VAR model, the dependent variable (e.g. symptom 'sadness', item 1) at time point *t* (e.g. session 2) is regressed on the lagged *t*−1 (e.g. session 1) versions of the independent variables (Box *et al.* 1994; Walls & Schafer, 2006).[4] The independent variables in this study are all the symptoms of the BDI-II, measured at the previous time point (in this case the previous session). To account for differences between patients, all regression coefficients were assumed to be normally distributed at the population level. As a consequence, we obtained a multilevel model consisting of fixed (average) and random (individual) effects.[5] Each BDI-II symptom was used as a criterion variable once, which means that 21 multilevel VAR models were estimated.

In order to estimate a multilevel VAR model, data need to be stationary. An implication of this assumption is that that the variables will fluctuate around the same mean over time (Lütkepohl, 2005). Since BDI-II symptoms decreased over the course of treatment (Lemmens *et al.* unpublished observations), the means changed significantly, which indicates a non-stationary process. For this reason, a linear trend in the multilevel VAR model was included, making the data trend stationary (Hamaker & Dolan, 2009). This implies that the short-term dynamics or the session-to-session fluctuations of the symptoms (as represented by the network) and the decrease of symptoms across the sessions (as represented by the linear trend) are modelled separately. Therefore, change in the short-term dynamics is in principle unrelated to change in the *mean level* of the BDI-II symptoms. Note further that stationarity also implies the assumption that the effects of symptoms on other symptoms are stable across time.

In order to obtain the BDI-II network, the estimated fixed effects of the multilevel VAR analyses were used (Snijders & Bosker, 2012). Fixed effects represent the average connection strengths of the arrows in the network among the 21 symptoms and indicate whether

the symptoms are positively or negatively related to each other. The fixed effects represent either autoregressive effects (self-loops) or cross-regressive effects (connections between different variables) in the network. Note that the network only represents the dynamic relations between the symptoms (the slopes of the multilevel VAR model) and not the mean scores (the intercepts of the multilevel VAR model) of the symptoms.

The estimated fixed effects or connections of the network resulted in a directed weighted network structure of the BDI-II, which was visualized using qgraph (Epskamp *et al.* 2012), a package for the statistical programming language R. Arrows or connections in the network represent more than mere associations between symptoms: because symptoms are measured over time, the connections can be viewed as an approximation of causality, resembling Granger causality (Granger, 1969; Tschacher *et al.* 2012). The network analyses were based on all the connections of the network. However, for reasons of clarity, we only visually present the strongest connections in the inferred network; that is, those connections which surpass the significance threshold (5%) using the false discovery rate (FDR) method (Benjamini & Hochberg, 1995; Benjamini *et al.* 2006). In the visually presented network, symptoms that are more strongly related to each other tend to be closer together in the figure (this is a result of the node placement algorithm; see Fruchterman & Reingold, 1991; Epskamp *et al.* 2012).

Since the current study included two different therapy groups (CT and IPT), it is possible that two different network structures give rise to the data. We tested this in two ways: first, we fitted a model with the two networks separately and cross-correlated their estimated network links. Second, we compared a model in which we included two networks with a model which had one common network; for this purpose, we used Bayesian Information Criterion (BIC; Schwarz, 1978). The model with the lowest BIC is the preferred model.[6]

Centrality analysis

In the second analysis, the inferred network was further analysed by estimating the centrality of the BDI-II symptoms. In a centrality analysis, one can determine the relative importance or influence of a symptom in the network. We performed three types of centrality analyses: outdegree, indegree and betweenness centrality (see Opsahl *et al.* 2010).[7] Outdegree centrality indicates how many outgoing arrows or how much information a symptom sends to other symptoms it is directly connected to. In the

same way, indegree centrality indicates how many incoming arrows a symptom receives from the directly connected symptoms. Betweenness centrality takes into account both the direct and indirect connections of a symptom. A symptom with a high betweenness centrality is a symptom located on many paths between other symptoms and thus is a symptom through which the information in the network has to pass often. Therefore, a symptom with a high betweenness centrality is important in funneling the information flow or the symptom spread in the network.
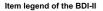
Community structure analysis

As a third analysis, we performed a community structure analysis. In complex networks, new structures of clusters can often be found. An example of such a cluster is a *community*, in which groups of symptoms are densely interconnected among each other, but sparsely connected to the overall network. We used the Walktrap algorithm, which is suited for weighted networks (Pons & Latapy, 2005). This algorithm does not take directions of the arrows into account, so we summed the connection strengths (arrows) between two symptoms to have an appropriate undirected network suitable for analysis. The Walktrap algorithm uses random walks on the network to find communities or densely interconnected symptoms. The algorithm reveals how many groups can be found and also to which group a symptom of the network belongs to. All the analyses were performed with statistical software R (R Core Team, 2014).

Results

The BDI-II network

Fig. 1 shows the inferred network of the dynamics between the 21 BDI-II symptoms. The analysis of cross-correlations and the model-fitting approach using BIC indicated that the network structure did not differ across the two therapy groups ($r=0.86$, $p<0.0001$; one network: BIC=77367.65 *v.* two networks: BIC=80574.09). Therefore, only one network was needed, representing both treatment groups. From the figure, it is evident that the strongest connections between symptoms are all positive in sign. Thus, when a symptom score increases, it is likely that other symptom scores also increase the next session, leading to an increase in the severity of symptoms in general. For example, if a participant reports feelings of guilt ('guilty feelings', item 5) in one session, that participant is more likely to report feelings of failure about the past ('past failure', item 3) the next session. The strength of the relationship between symptoms translates into the thickness of the arrows in the figure: the stronger the

Item legend of the BDI-II
1. Sadness
2. Pessimism
3. Past Failure
4. Loss of Pleasure
5. Guilty Feelings
6. Punishment Feelings
7. Self-Dislike
8. Self-Criticalness
9. Suicidal Thoughts
10. Crying
11. Agitation
12. Loss of Interest
13. Indecisiveness
14. Worthlessness
15. Loss of Energy
16. Changes in Sleeping
17. Irritability
18. Changes in Appetite
19. Concentration Difficulty
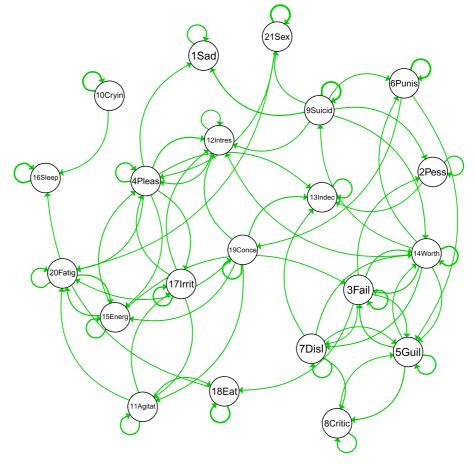20. Tiredness
21. Loss of Interest in Sex

Fig. 1. The BDI-II network. In this network, the connections between the 21 symptoms that surpass the significance threshold are visualized. Because of multiple hypothesis testing, we do not use the traditional 0.05 cut-off for *p* values as the standard (which would inflate the number of unimportant links to be visualized); instead, we control the false discovery rate (FDR) at 5% (Benjamini & Hochberg, 1995; Benjamini *et al.* 2006). Here, the 75 connections that pass the FDR threshold are visualized.

symptoms are related the thicker the arrow between two symptoms, and the closer the symptoms tend to be together in the figure. This is expressed in, for example, the placement of the symptoms 'past failure' (item 3) and 'worthlessness' (item 14).
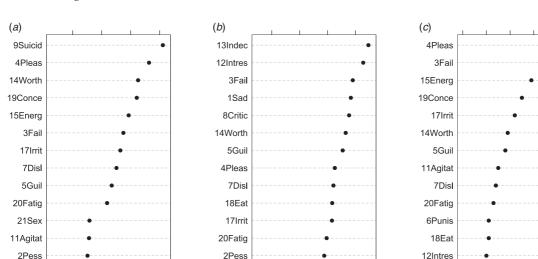
Apart from the connections between the symptoms, self-loops can contain important information. For example, the self-loop of the symptom 'loss of interest in sex' (item 21) is clearly the strongest connection of the network, meaning that when a participant reports loss of interest in sex one session, he or she is highly likely to report this in the next session as well. Furthermore, self-sustaining loops are apparent in the network. For example, 'worthlessness' (item 14) and 'guilty feelings' (item 5) seem to mutually influence each other. It should be mentioned that there are negative connections in the complete network as well. However, since these are rather weak, they did not pass the threshold for visualization in Fig. 1. Note, however, that all connections are taken into account in the further analyses.[8,9]

Centrality analysis

Fig. 2 presents the results of the centrality analysis. Fig. 2*a* indicates that the symptom 'loss of pleasure' (item 4) has one of the highest outdegrees, meaning that when one reports loss of pleasure in one session, it is likely that one will also report an increase in other symptoms in the next session. This is in contrast to, for instance, the symptom 'changes in sleeping patterns' (item 16), which is less likely to directly affect other symptoms the next session.

Fig. 2*b* indicates that the symptoms 'indecisiveness' (item 13), 'loss of interest' (item 12), 'past failure' (item 3) and 'sadness' (item 1) feature higher indegrees and thus receive a lot of information from other symptoms. This is in contrast to 'suicidal thoughts' (item 9): this symptom is unlikely to be influenced by other symptoms, and is more likely to influence other symptoms (see also Fig. 2*a*).

Fig. 2*c* indicates that the symptoms 'loss of pleasure' (item 4) and 'past failure' (item 3) feature the highest

Fig. 2. (*a*) Outdegree, (*b*) indegree and (*c*) betweenness centrality for all BDI-II symptoms. The black dots are the model-based estimates of outdegree, indegree, and betweenness centrality. The higher the centrality index score the more central the symptom is in the network.

betweenness centralities, but they also have one of the highest outdegree ('loss of pleasure') and indegree ('past failure') centrality scores, respectively. Thus, the symptoms 'loss of pleasure' and 'past failure' are important in funneling the activation flow or symptom spread in the network.

Community structure of the BDI-II network

The community structure analysis using the Walktrap algorithm indicated a two-cluster solution (see Fig. 3).[10] This community structure means that symptoms in one cluster are more densely interconnected among themselves and more sparsely connected to symptoms in another cluster. The green cluster in Fig. 3 consists of the symptoms 'guilty feelings' (item 5), 'past failure' (item 3), 'self-dislike' (item 7), 'self-criticalness' (item 8), 'worthlessness' (item 14), 'punishment feelings' (item 6) and 'pessimism' (item 2), which are often described as cognitive symptoms. Items in the yellow cluster mainly consist of physical and affective symptoms of depression that appear related to loss of energy and pleasure.

Discussion

In this paper, we derived for the first time a network that represents the session-to-session dynamics of one of the most widely used and empirically validated self-report measures for assessing the severity of depression: the BDI-II (Beck *et al.* 1996*b*). Results indicate that, in this network, all BDI-II symptoms are directly or indirectly connected. In addition, the strongest connections between the symptoms are uniformly positive, indicating that, in general, when a symptom changes in severity, other symptoms tend to change in the same direction. This pattern of symptom dynamics is independent of the overall decrease in symptom scores as this trend was modelled separately; hence, the pattern does not reflect the common influence of treatment or recovery. In addition, for each symptom–symptom relation, we controlled for the effect of all other symptoms in the network. Thus, the evidence from this study points to the conclusion that direct effects among symptoms of the BDI-II are prevalent, and in fact connect all symptoms assessed in the questionnaire. In other words, symptoms of depression form a network of direct interactions.
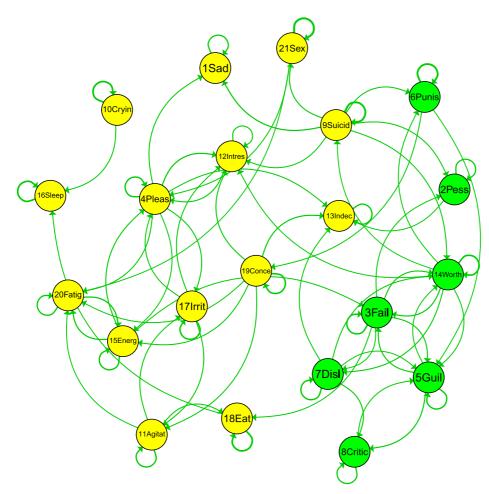
Fig. 3. Community structure of the BDI-II network with the two clusters indicated by two different colours.

Centrality analyses of the symptoms suggested that some symptoms are likely to have a larger influence on the symptom spread than other symptoms. As one may expect based on, e.g. the DSM-IV, the symptom 'loss of pleasure' (item 4) is one of the most central items in the symptom network and thus has a relatively large effect on the enduring of depressive symptoms in general. Somewhat more surprisingly, the symptoms 'sadness' (item 1) and 'loss of interest' (item 12) have a high indegree centrality, but quite a low outdegree and betweenness centrality, which means that they tend to increase in severity as other symptoms become more severe, but do not play a large role in funneling the symptom spread themselves. Thus, these symptoms may serve a mainly reactive role in the short-term dynamics of depression. Additional studies are needed to confirm these results, preferably engaging different depression questionnaires, such as the HAMD.

Based on theory, one may expect a difference in symptom dynamics for subjects receiving CT and IPT because both treatments are assumed to work through different mechanisms. We did not observe such differences. A potential explanation for our findings could be that the dynamics between symptoms are similar when the treatments that are being compared are equally effective in reducing pathology, a fact that has been well established for CT and IPT for depression (Cuijpers *et al.* 2008; Hollon & Ponniah, 2010). Alternatively, it might be the case that differences between CT and IPT actually do exist, but that we did not capture them in the current study because the BDI-II is, due to its design, insensitive for the differences between the two treatments. For example, even though the BDI-II includes items on several cognitive components (key elements of CT), items referring to problems in the interpersonal domain (core of IPT) are lacking. Further research involving other questionnaires is necessary to indicate if there are differences in symptom dynamics between therapies. A final possibility is that the difference between CT and IPT does not lie in symptom–symptom interaction, as studied in this paper, but in differences that arise in, e.g. stepwise changes in symptomatology. In this case, therapy

effects might be detected in the way symptoms decrease or increase from one time point to another. Models that may be used to analyse such differences, while accounting for the network of symptom-symptom interaction, are currently unavailable; however, nonlinear statistical network inference techniques that may be used to model such processes are within reach, and could be used to investigate this issue in the future.

In the present study, community analyses revealed two groups of symptoms. The result appears to accommodate emerging evidence from the biomedical literature, which points to two types of depression: melancholia and atypical depression (Lamers *et al.* 2010, 2012). The current community clusters resemble these different depression types, as the green cluster in Fig. 3 has similarities to melancholic type, whereas the yellow cluster resembles atypical depression. It is also interesting to note that the community structure result, based on multiple time points, is similar to the two-factor solution of the BDI-II, based on pooling across subjects at one time point (as found in e.g. Beck *et al.* 1996*b*; see also Steer *et al.* 1999; Arnau *et al.* 2001). Except for 'suicidal thoughts', all other symptoms in the green cluster of Fig. 3 are the same as in the cognitive dimension of the two-factor solution of the BDI-II, whereas the yellow cluster could be interpreted as the somatic-affective or non-cognitive dimension. Although it is a good sign that the results we find are consistent with what one typically finds using factor analysis, our approach leads to a different way of thinking, different strategies for intervention, and to very different conclusions. In the latent variable approach, there are just two clusters of symptoms, which is a static result. In the network view, the result concerns the communication between symptoms that is denser within the cluster than with symptoms that are not in the cluster, leading to new hypotheses on how interventions should be operationalized, namely focusing on the interaction between symptoms. Thus, the existence of such patterns of influence is not a replication of the results of factor analysis on individual differences; rather, it may be seen as a potential explanation for these results (Wichers, 2014).

Several findings of this paper suggest further research. One important issue is how our results, which only involved participants with a diagnosis of depression, compare to results from unaffected individuals. For example, it is important to investigate whether a similar network characterizes healthy individuals. One hypothesis would be that there are no distinct symptom clusters in healthy subjects, but that instead all symptoms are similarly (and weakly) connected. Such a network would be more resilient, since activation would not spread as easily, and it would be less likely to 'get stuck' in a cluster of symptoms. Another important topic for future research involves the difficult question of how to relate different time scales (Boker *et al.* 2009). This is because the symptoms that characterize depression are likely to influence each other in different time windows. For example, sleep problems are likely to exert effects in a pattern of a day-to-day variation, whereas mood states are much quicker and may affect each other within minutes. The question of how the dynamics of these different time scales interact with each other is, in our view, one of the main puzzles to be solved in the study of symptom dynamics.

Regarding clinical practice, the relevance of the methodology and results of our approach may lie in opportunities to determine symptom centrality. For example, network analyses may be used to indicate which symptoms should be targeted first, and in this sense may help in setting up treatment strategies. Ideally, such analyses should be based on person-specific analyses (cf. Molenaar & Campbell, 2009). Unfortunately, at the moment such analyses are not computationally feasible for large networks of 21 symptoms. However, future development of the multilevel VAR method, combined with a higher frequency of within subject assessment, should make it possible to take this procedure a step further, which may eventually lead to person-specific therapeutic interventions. Information about person-specific network centrality would not necessarily require pre-treatment assessment, and the high frequency assessment could be informative at any point, even if started during therapy. For example, if a centrality analysis of an individual network reveals that for that specific person 'loss of pleasure' is the most central symptom, therapy that intervenes on this symptom would be more effective than treatment that intervenes on non-central symptoms; for other persons, different interventions may be preferable. In a similar vein, one could hypothesize that if 'suicidal thoughts' is the most central symptom for a given person, this may signal acute need for care. Furthermore, since 'suicidal thoughts' has a high outdegree, and is thus likely to trigger other symptoms, but a low indegree, and is thus not likely to be influenced by the other symptoms, interventions should be directly targeted at this symptom. Given the increased opportunities for assessing highly intensive time series within individuals, person-specific treatment protocols based on networks of symptom dynamics are rapidly becoming a realistic possibility. Thus, the network perspective is a promising new research field, which can give guidance to research on depression and to psychological research in general.

Acknowledgements

Declaration of Interest

None.

Notes

[1] Note that short-term dynamics refers to the dynamics between time points that are close to one another (e.g. time point $t-1$ and t). This in contrast to looking at changes in average values (mean levels), which can also be seen as a long-term dynamics. However, we do not use the latter term, since the term 'change in mean level' is more intuitive.

[2] Analyses indicated that there were no differences in demographic and clinical characteristics between subjects that dropped out and those who finished therapy (at least 12 sessions).

[3] Note that we deviate in this paper slightly from the procedure as proposed by Bringmann *et al.* (2013). With 21 items, it is not computationally possible to include all 21 random effects in the multilevel VAR model simultaneously. Instead, we included only five random effects (including the autoregressive coefficient and the intercept) at the same time in a stepwise manner. Simulations indicated that the fixed effects could still be estimated precisely with this number of subjects and time points, which means that this is a feasible approach for estimating the current average network.

[4] Theoretically further lags are also possible. For example, a lag 2 model would indicate how symptoms are related to all symptoms experienced two sessions and one session ago. However, model comparison indicated that lag 1 was a more likely model than a lag 2 model (BIC lag 1: 71162, BIC lag 2: 71539).

[5] Simulations (not reported here) have indicated that because it is computationally not possible to include all 21 random effects at once in the multilevel VAR model, the variance components (random effects variances) cannot be estimated accurately enough. For this reason, they will not be discussed further in the paper. The random effects should not be left out of the model though, because their inclusion leads to a more precise estimate of the fixed effects.

[6] The BIC was calculated by taking the average of the BICs of the separate univariate models.

[7] Since we want to estimate the centrality between the symptoms, self-loops are not taken into account in the centrality analyses. However, in all other analyses self-loops are taken into account.

[8] We also confirmed that the connections in the network and thus the relationships between symptoms are not driven by differential variability. Standardizing the data per patient and per symptom led to a network that was highly similar to the original network; the correlation between parameters in the original and standardized network was 0.99. As a result, the conclusions of this paper are robust with respect to standardization of the data and are unlikely to reflect differential symptom variability.

[9] Proportional odds logistic regression (POLR), which is a regression model for ordinal response variables, also showed highly similar results; the correlation between parameters in the original and POLR network was 0.96 and led to similar centrality and community cluster results.

[10] A hierarchical cluster analysis on the sum of the weighted links gave highly similar results.

References

APA (2000). *Diagnostic and Statistical Manual of Mental Disorders*, 4th edn, Text Revision. American Psychiatric Association: Washington, DC.

Arnau RC, Meagher MW, Norris MP, Bramson R (2001). Psychometric evaluation of the Beck Depression Inventory-II with primary care medical patients. *Health Psychology* **20**, 112–119.

Beck AT (1964). Thinking and depression: II. Theory and therapy. *Archives of General Psychiatry* **10**, 561–571.

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

Beck AT, Steer RA, Ball R, Ranieri W (1996*a*). Comparison of Beck Depression Inventories -IA and -II in psychiatric outpatients. *Journal of Personality Assessment* **67**, 588–597.

Beck AT, Steer RA, Brown GK (1996*b*) *Manual for the Beck Depression Inventory*, 2nd edn. Psychological Corporation: San Antonio, TX.

Beck AT, Ward CH, Mendelson M, Mock J, Erbaugh J (1961). An inventory for measuring depression. *Archives of General Psychiatry* **4**, 561–571.

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B Methodological* **57**, 289–300.

Benjamini Y, Krieger AM, Yekutieli D (2006). Adaptive linear step-up procedures that control the false discovery rate. *Biometrika* **93**, 491–507.

Bhar SS, Gelfand LA, Schmid SP, Gallop R, DeRubeis RJ, Hollon SD, Amsterdam JD, Shelton RC, Beck AT (2008). Sequence of improvement in depressive symptoms across cognitive therapy and pharmacotherapy. *Journal of Affective Disorders* **110**, 161–166.

Boccaletti S, Latora V, Moreno Y, Chavez M, Hwang D-U (2006). Complex networks: structure and dynamics. *Physics Reports* **424**, 175–308.

Boker SM, Molenaar P, Nesselroade JR (2009). Issues in intraindividual variability: individual differences in equilibria and dynamics over multiple time scales. *Psychology and Aging* **24**, 858–862.

Borsboom D (2008). Psychometric perspectives on diagnostic systems. *Journal of Clinical Psychology* **64**, 1089–1108.

Box JEP, Jenkins GM, Reinsel GC (1994). *Time Series Analysis: Forecasting and Control*. Prentice Hall: Englewood Cliffs, NJ.

Bringmann LF, Vissers N, Wichers M, Geschwind N, Kuppens P, Peeters F, Borsboom D, Tuerlinckx F (2013). A network approach to psychopathology: new insights into clinical longitudinal data. *PLoS ONE* **8**, e60188.

Cramer AOJ, Borsboom D, Aggen SH, Kendler KS (2012). The pathoplasticity of dysphoric episodes: differential impact of stressful life events on the patterns of depressive symptom inter-correlations. *Psychological Medicine* **42**, 957–965.

Cramer AOJ, Waldorp LJ, van der Maas HLJ, Borsboom D (2010). Comorbidity: a network perspective. *Behavioral and Brain Sciences* **33**, 137–193.

Cuijpers P, Geraedts AS, van Oppen P, Andersson G, Markowitz JC, van Straten A (2011). Interpersonal psychotherapy for depression: a meta-analysis. *American Journal of Psychiatry* **168**, 581–592.

Cuijpers P, van Straten A, Andersson G, van Oppen P (2008). Psychotherapy for depression in adults: a meta-analysis of comparative outcome studies. *Journal of Consulting and Clinical Psychology* **76**, 909–922.

Epskamp S, Cramer AOJ, Waldorp LJ, Schmittmann VD, Borsboom D (2012). Qgraph: network representations of relationships in data. R package version 0.4.10 (http://CRAN.R-project.org/package=qgraph).

Fournier JC, DeRubeis RJ, Hollon SD, Gallop R, Shelton RC, Amsterdam JD (2013). Differential change in specific depressive symptoms during antidepressant medication or cognitive therapy. *Behaviour Research and Therapy* **51**, 392–398.

Fried EI, Nesse RM, Zivin K, Guille C, Sen S (2013). Depression is more than the sum score of its parts: individual DSM symptoms have different risk factors. *Psychological Medicine*. Published online: 2 December 2013. doi:10.1017/S0033291713002900.

Fruchterman TMJ, Reingold EM (1991). Graph drawing by force-directed placement. *Software – Practice and Experience* **21**, 1129–1164.

Girvan M, Newman M (2002). Community structure in social and biological networks. *Proceedings of the National Academy of Sciences* **99**, 7821–7826.

Granger CWJ (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica* **37**, 424–438.

Hamaker EL, Dolan CV (2009). Idiographic data analysis: quantitative methods – from simple to advanced. In *Dynamic Process Methodology in the Social and Developmental Sciences*, pp. 191–216. Springer: New York.

Hamilton M (1960). A rating scale for depression. *Journal of Neurology, Neurosurgery, and Psychiatry* **23**, 56–62.

Hardeveld F, Spijker J, De Graaf R, Nolen WA, Beekman ATF (2010). Prevalence and predictors of recurrence of major depressive disorder in the adult population. *Acta Psychiatrica Scandinavica* **122**, 184–191.

Hollon SD, Ponniah K (2010). A review of empirically supported psychological therapies for mood disorders in adults. *Depression and Anxiety* **27**, 891–932.

Hollon SD, Thase ME, Markowitz JC (2002). Treatment and prevention of depression. *Psychological Science in the Public Interest* **3**, 39–77.

Kendler KS, Zachar P, Craver C (2011). What kinds of things are psychiatric disorders? *Psychological Medicine* **41**, 1143–1150.

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

Klerman GL, Weissman MM, Rounsaville BJ, Chevron ES (1984). *Interpersonal Psychotherapy for Depression*. Basic Books: New York.

Lamers F, de Jonge P, Nolen WA, Smit JH, Zitman FG, Beekman AT, Penninx BW (2010). Identifying depressive subtypes in a large cohort study: results from the Netherlands Study of Depression and Anxiety (NESDA). *Journal of Clinical Psychiatry* **71**, 1582–1589.

Lamers F, Rhebergen D, Merikangas KR, de Jonge P, Beekman ATF, Penninx BWJH (2012). Stability and transitions of depressive subtypes over a 2-year follow-up. *Psychological Medicine* **17**, 1–11.

Lemmens LH, Arntz A, Peeters FP, Hollon SD, Roefs A, Huibers MJ (2011). Effectiveness, relapse prevention and mechanisms of change of cognitive therapy *vs.* interpersonal therapy for depression: study protocol for a randomised controlled trial. *Trials* **12**, 150.

Lütkepohl H (2005). *New Introduction to Multiple Time Series Analysis*. Springer: Berlin.

Molenaar PC, Campbell CG (2009). The new person-specific paradigm in psychology. *Current Directions in Psychological Science* **18**, 112–117.

Opsahl T, Agneessens F, Skvoretz F (2010). Node centrality in weighted networks: generalizing degree and shortest paths. *Social Networks* **32**, 245–251.

Pons P, Latapy M (2005). Computing communities in large networks using random walks. In *Computer and Information Sciences-ISCIS*, pp. 284–293. Springer: Berlin.

R Core Team (2014). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing: Vienna, Austria (http://www.R-project.org/).

Rush AJ, Kovacs M, Beck AT, Weissenburger J, Hollon SD (1981). Differential effects of cognitive therapy and pharmacotherapy on depressive symptoms. *Journal of Affective Disorders* **3**, 221–229.

Schwarz G (1978). Estimating the dimension of a model. *Annals of Statistics* **6**, 461–464.

Snijders T, Bosker R (2012). *Multilevel Analysis: an Introduction to Basic and Advanced Multilevel Modeling.* Sage Publications: London.

Steer RA, Ball R, Ranieri WF, Beck AT (1999). Dimensions of the Beck Depression Inventory-II in clinically depressed outpatients. *Journal of Clinical Psychology* **55**, 117–128.

Stewart JG, Harkness KL (2012). Symptom specificity in the acute treatment of major depressive disorder: a re-analysis of the treatment of depression collaborative research program. *Journal of Affective Disorders* **137**, 87–97.

Tschacher W, Zorn P, Ramseyer F (2012). Change mechanisms of schema-centered group psychotherapy with personality disorder patients. *PLoS ONE* **7**, e39687.

Van der Does AJW (2002). *Manual of the Dutch Version of the Beck Depression Inventory (BDI-II-NL).* Harcourt Publishers: Amsterdam.

Walls TA, Schafer JL (2006). *Models for Intensive Longitudinal Data.* Oxford University Press: Oxford.

Wichers M (2014). The dynamic nature of depression: a new micro-level perspective of mental disorder that meets current challenges. *Psychological Medicine* **44**, 1349–1360.

WHO (2001). *The World Health Report. Mental Health: New Understanding, New Hope.* World Health Organization: Geneva.

WHO (2008). *ICD-10: International Statistical Classification of Diseases and Related Health Problems*, 10th Rev. edn. Author: New York, NY.