

Research paper

## The centrality of DSM and non-DSM depressive symptoms in Han Chinese women with major depression



Kenneth S. Kendler<sup>a,\*</sup>, Steven H. Aggen<sup>a</sup>, Jonathan Flint<sup>b</sup>, Denny Borsboom<sup>c</sup>, Eiko I. Fried<sup>c</sup>

<sup>a</sup> Virginia Institute of Psychiatric and Behavioral Genetics and the Departments of Psychiatry, Virginia Commonwealth University School of Medicine, Richmond, VA, USA

<sup>b</sup> Department of Psychiatry, University of California at Los Angeles, Los Angeles, CA, USA

<sup>c</sup> Department of Psychological Methods, University of Amsterdam, USA

### ARTICLE INFO

**Keywords:**

Major depression

Diagnostic and Statistical Manual 3rd Edition

Network analysis

### ABSTRACT

**Introduction:** We compared DSM-IV criteria for major depression (MD) with clinically selected non-DSM criteria in their ability to represent clinical features of depression.

**Method:** We conducted network analyses of 19 DSM and non-DSM symptoms of MD assessed at personal interview in 5952 Han Chinese women meeting DSM-IV criteria for recurrent MD. We estimated an Ising model (the state-of-the-art network model for binary data), compared the centrality (interconnectedness) of DSM-IV and non-DSM symptoms, and investigated the community structure (symptoms strongly clustered together).

**Results:** The DSM and non-DSM criteria were intermingled within the same symptom network. In both the DSM-IV and non-DSM criteria sets, some symptoms were central (highly interconnected) while others were more peripheral. The mean centrality of the DSM and non-DSM criteria sets did not significantly differ. In at least two cases, non-DSM criteria were more central than symptomatically related DSM criteria: lowered libido vs. sleep and appetite changes, and hopelessness versus worthlessness. The overall network had three sub-clusters reflecting neurovegetative/mood symptoms, cognitive changes and anxiety/irritability.

**Limitations:** The sample were severely ill Han Chinese females limiting generalizability.

**Conclusions:** Consistent with prior historical reviews, our results suggest that the DSM-IV criteria for MD reflect one possible sub-set of a larger pool of plausible depressive symptoms and signs. While the DSM criteria on average perform well, they are not unique and may not be optimal in their ability to describe the depressive syndrome.

### 1. Introduction

The history of the symptoms and signs used as diagnostic criteria for major depression (MD) in DSM-III and subsequent DSM editions is relatively well understood ((Kendler et al., 2010) Table 1). They derive, with minimal changes, from those proposed for the Research Diagnostic Criteria (Spitzer et al., 1975), which in turn were based, with modest modifications, on those included in the Feighner Criteria (Feighner et al., 1972). The Feighner criteria for MD were themselves adapted from an earlier set of items proposed by Cassidy et al. (1957) who cite, as one key source, a set of criteria for MD proposed previously by Stone and Burris (1950). Some differences across these criteria were noteworthy. For example, Cassidy et al. (1957) included slowed thinking, decreased libido and constipation, none of which were included in DSM-III. DSM-III added worthlessness, a symptom not present in the earlier diagnostic formulations, and added appetite/weight gain, not present in either Cassidy et al. (1957) or Stone and Burris (1950).

A recent review provided a broader historical context within which to view the DSM criteria for MD (Kendler, 2016). Examining textbook descriptions of the depressive syndrome from 1900 to 1960, a good but imperfect correspondence was seen between symptoms and signs noted by historical experts and those incorporated into the recent DSM editions. Of the 18 depressive symptoms and signs frequently noted by these textbook authors, 10 were well covered by DSM MD criteria, two were partly covered and six were entirely absent (Kendler, 2016). For example, the historical experts noted that symptoms of anxiety were commonly present in depression but these were not included in any modern MD criteria. In describing the common cognitive changes in depression, the textbook authors noted a rather wide range of symptoms including hopelessness, pessimism and feelings of inadequacy, symptoms not entirely captured by the relevant single DSM criterion which assesses guilt and feelings of worthlessness. These results were recently extended further back in time to the critical period between 1880 and 1900 where expert descriptions of the depressive syndrome

\* Correspondence to: Virginia Institute for Psychiatric and Behavioral Genetics of VCU, Box 980126, Richmond, VA 23298-0126, USA.  
E-mail address: kenneth.kendler@vcuhealth.org (K.S. Kendler).

**Table 1**

Characteristics of the DSM-IV and non-DSM Symptoms Obtained at Personal Interview in the CONVERGE Sample (n = 6008).

Criterion #	Label	Proportion positive	Std dev	N Missing
1	A1: sad mood	0.995	0.073	0
2	A2: ↓interest	0.989	0.107	0
3	A3: Δ wt, app	0.907	0.290	40
4	A4: Δ sleep	0.956	0.205	49
5	A5: motor Δs	0.910	0.287	41
6	A6: fatigue	0.935	0.246	40
7	A7: worthless	0.904	0.295	41
8	A8: diff conc	0.977	0.150	40
9	A9: suicide	0.764	0.425	41
10	↓self-esteem	0.843	0.364	40
11	↓confidence	0.863	0.342	40
12	distinct quality	0.930	0.255	50
13	worse in AM	0.622	0.485	47
14	↓libido	0.890	0.313	49
15	unreactive mood	0.866	0.341	49
16	irritable/angry	0.747	0.435	48
17	hopeless	0.806	0.396	50
18	crying	0.674	0.469	49
19	helpless	0.894	0.307	49
20	nervous	0.892	0.310	50

closely resembled those found from 1900 to 1960 (Kendler, 2017b).

These historical inquiries suggest that the specific criteria chosen for MD for the DSM-III and subsequent DSM editions reflect one subset of a broader number of plausible criteria that could have been chosen. This viewpoint is supported by evidence that common rating scales for depression differ widely in the symptoms they assess (Fried, 2016). From this perspective, it naturally becomes of interest to examine how the DSM criteria for MD might compare to a set of other plausible symptoms of depression not included in the DSM. Are there distinctive features which differentiate DSM-criteria for MD from these other depressive symptoms? Are the DSM-criteria more centrally placed in this structure of depression than are credible non-DSM depressive symptoms?

To address this question, we utilize a network approach with which we quantify – via the concept of centrality (Opsahl et al., 2010) – how closely interconnected each individual criterion is with all the other symptoms in the network. Specifically, we apply network analysis to 8 of the 9 DSM depressive criteria and 11 other depressive symptoms chosen for their research and clinical value. All these criteria were assessed at personal interview in Han Chinese women, ascertained in psychiatric treatment facilities, who met DSM-IV criterion for recurrent MD. We first describe the network formed by these 19 putative criteria and determine the degree to which the DSM and non-DSM criteria are part of a single network. Second, we explore the connectivity structure of these depressive symptoms as revealed by our network analysis to determine if the DSM criteria are more central to the network than are the non-DSM symptoms.

## 2. Methods

### 2.1. Sample

The analyses here reported were based on a total of 6008 female cases of MD recruited as part of the CONVERGE (China, Oxford, and VCU Experimental Research on Genetic Epidemiology) study from 57 mental health centers and psychiatric departments of general medical hospitals in 45 cities in 23 provinces in China. The primary focus of CONVERGE was a molecular genetic study of MD (CONVERGE consortium, 2015). Given evidence that the genetic effects on MD are different in the sexes (Kendler et al., 2001), we included only female participants with four Han Chinese grandparents. Cases were excluded if they had a pre-existing history of bipolar disorder, psychosis or

mental retardation. Cases were aged between 30 and 60, had two or more episodes of MD meeting DSM-IV criteria (American Psychiatric Association, 1994) with the first episode occurring between 14 and 50, and had not abused drugs or alcohol before their first depressive episode.

All subjects were interviewed using a computerized assessment system. Interviewers were postgraduate medical students, junior psychiatrists or senior nurses, trained by the CONVERGE team for a minimum of one week. The study protocol was approved centrally by the Ethical Review Board of Oxford University and the ethics committee in the participating hospitals in China.

The diagnosis of MD was established with the Composite International Diagnostic Interview (CIDI) (WHO lifetime version 2.1; Chinese version), which operationalized DSM-IV criteria for MD (World Health Organization, 1990). The interview was originally translated into Mandarin by a team of psychiatrists in Shanghai Mental Health Center, with the translation reviewed and modified by members of the CONVERGE team.

From his clinical and research experience, one of us (KSK) added a range of additional items to the depression section of our CONVERGE interview with the goal of using these items to help further in the characterization of the depression syndrome and its relationship with genetic and environmental risk factors. These items came from a range of sources including Beck's work (Beck et al., 1980) and the DSM-IV criteria for melancholia, and their translation to Mandarin was performed and then checked. A number of items included were not used in these analyses because of missing data due to skip patterns which induced statistical dependencies. All items utilized were asked for all subjects included in the sample with low missingness resulting either from the rare refusal or inability to answer or software malfunctions. Only 56 of 6008 participants had missing data, resulting in a final dataset of n = 5952.

### 2.2. Statistical methods

Endorsement rates for sad mood were so high that it was not feasible to include this criterion in subsequent analyses. Therefore, we estimated the network structure among the remaining 8 DSM criterion for MD and 11 non-DSM symptoms, leading to a network with 19 nodes (the symptoms) and 171 potential connections among these symptoms. We followed the three steps for network analysis proposed recently: network estimation, network inference, and network robustness (Epskamp et al., 2017).

We estimated an Ising Model, the state-of-the-art network model in psychopathology research for binary data, (van Borkulo et al., 2014) which has two important characteristics. Symmetric pathways between the symptoms (called “edges” in the more technical network literature and “connections” here) are estimated as conditional dependence relations: an association between two symptoms means that they remain conditionally dependent after controlling for all other associations among the symptoms in the network. Conversely, if no edge emerges between two symptoms, they are conditionally independent after controlling for the associations among all other symptoms. One can think of the edges in the Ising Model as akin to partial correlations. Second, the network is regularized, is a statistical strategy that shrinks many connections in the network, and sets very small connections to exactly zero (Tibshirani, 1996). This results in a parsimonious (sparse) network structure that reduces the number of false positive connections and performs well to recover underlying network structures (van Borkulo et al., 2014).

In a second step, we investigated the degree centrality of each of the symptoms in the network (Opsahl et al., 2010), which we refer to simply as ‘centrality’ in the remainder of the paper. Centrality is defined here as the sum of the absolute values of retained connections for each symptom with all other symptoms. A central symptom is usually one with several strong connections, while a peripheral symptom usually

has no connections or only few that are mostly weak. Note that prior papers have often investigated further centrality metrics such as closeness or betweenness (Opsahl et al., 2010); we focus on degree centrality because it is more interesting for our research question, and because closeness and betweenness are often not reliably estimated in psychopathological networks (Epskamp et al., 2017). For node placement in the resulting network graph, we use the Fruchterman-Reingold algorithm (Fruchterman and Reingold, 1991) that iteratively computes the optimal layout so that symptoms with stronger and/or more connections are placed closer to each other; central nodes often end up in the center of the graph, nodes with low centrality in the periphery. To statistically compare whether DSM and non-DSM symptom sets differ in centrality, we used a permutation test suggested by Fried et al. (2016). This test randomly assigns symptoms to two groups 100,000 times, and estimates the degree to which these two groups differ in centrality, creating a distribution under the null-hypothesis. We can then evaluate whether the empirical centrality difference between DSM and non-DSM symptoms is more pronounced than would be expected under the null-hypothesis. We also investigated the community structure of the graph. A community is defined as a set of items that cluster more strongly amongst each other than with other items; the network perspective explains such communities as a result of increased mutual influences among symptoms in a given cluster. We used two state-of-the-art methods for community detection to ensure robustness of results: the walktrap algorithm (Pons and Latapy, 2005) and the spinglass algorithm (Reichardt and Bornholdt, 2006). Results were nearly identical, and we present the results of the walktrap algorithm in the paper and the results of the spinglass algorithm in the Appendix Figure 6.

In a third step, we investigated the accuracy and stability of the network model, using several bootstrapping procedures described in detail in a recent tutorial paper (Epskamp et al., 2017). These tests (a) reveal how accurately the connections are estimated by constructing a 95% confidence interval (CI) around them, and (b) show how stable centrality is estimated via the centrality-stability coefficient.

We estimated and generated graphical visualizations of the binary symptom network via the R-packages IsingFit (van Borkulo and Epskamp, 2016) and qgraph (Epskamp et al., 2012), and tested the robustness using the R-package bootnet (Epskamp and Fried, 2015). To increase the reproducibility of our results, we have made all code available online (<https://osf.io/2bj8q/>).

### 3. Results

Table 1 depicts the frequency of endorsement of the 20 putative depressive criteria assessed during the worst lifetime episode. These rates varied from a low of 62.2% for *worse in AM* to a high of 99.6% for *sad mood*. The DSM criteria had, on average, higher endorsement rates than did the non-DSM criteria. A tetrachoric correlation matrix for these 20 criteria, along with the adjacency matrix (i.e. the numerical value of all connections) of the network depicted in Fig. 1, is presented in Appendix Table 1.

The estimated network shown in Fig. 1 displays the DSM and non-DSM criteria as grey and white circles, respectively. Three features of the network are noteworthy and will be subsequently evaluated statistically. First, the DSM and non-DSM criteria sets are part of the same intermingled network of depressive symptoms – that is, they are substantially interrelated with each other. Second, the network has a clinically meaningful sub-structure. Neurovegetative and mood symptoms cluster at the top of the network, anxious/irritable symptoms at the lower left and cognitive features of depression on the lower right. Third, some of the DSM criteria, especially *psychomotor symptoms* (#5), *worthlessness* (#7) and *difficulty concentrating* (#8) appear relatively highly inter-connected with other criteria in the network. However, other DSM criteria, such as *decreased interest* (#2) and *weight/appetite changes* (#3), are more poorly inter-connected. We see a similar variability with the non-DSM criteria which include two peripheral

items (*distinct quality* [#12] and *worse in AM* [#13]) and several which appear highly inter-connected including *unreactive mood* (#15) *decreased libido* (#14), *hopelessness* (#17) and *decreased self-esteem* (#10).

Fig. 2 summarizes the strength of inter-connectedness – that is centrality – of each of the 19 criteria included in the network. The results align with our “eye-ball” summary of the network graph. *Psychomotor changes* (DSM), *hopelessness* (non-DSM) and *decreased self-confidence* (non-DSM) are the most central criteria while the least central are *distinct quality* and *worse in AM* (both non-DSM) and *loss of interest* (DSM). These results permitted a formal comparison of the interconnectedness of the DSM and non-DSM criteria which had mean centralities of 3.58 and 3.35, respectively [SDs 1.56 and 1.73]. They do not significantly differ from each other either by a Welch Two Sample t-test [ $t = 0.3$ ,  $df = 16.01$ ,  $p = 0.77$ ] or a permutation test:  $p = 0.76$ . As seen in the Appendix Figs. 1–4, robustness analyses showed that both the strength of the connections and the centrality of the individual symptoms were quite reliably estimated as might be expected given the large sample size.

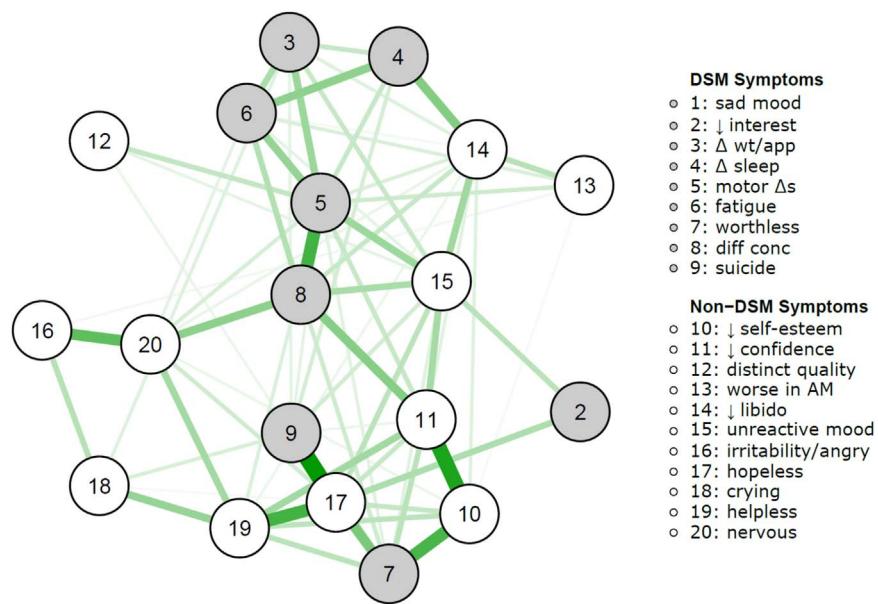
Fig. 3 presents the results of the community detection analysis using the walktrap algorithm. (Results presented in the Appendix Figure 6 were very similar using a different method [the spinglass algorithm] to determine the network structure). We see three clinically meaningful clusters. The largest, cluster A depicted in red in the figure, included 9 symptoms that reflected mood (3 items), somatic/neurovegetative symptoms (4 items) and difficulty with concentration (1 item). The second largest cluster B depicted in green, included 7 symptoms which all reflected cognitive changes associated with depression. The smallest cluster C included three items which reflected anxiety/irritability. While cluster A was made up of nearly equal numbers of DSM and non-DSM criteria, cluster B was predominantly and cluster C entirely made up of non-DSM criteria. Interestingly, helplessness was the key bridging criterion between clusters B and C while difficulty with concentration seemed the most important transitional symptom between clusters A and B and A and C.

### 4. Discussion

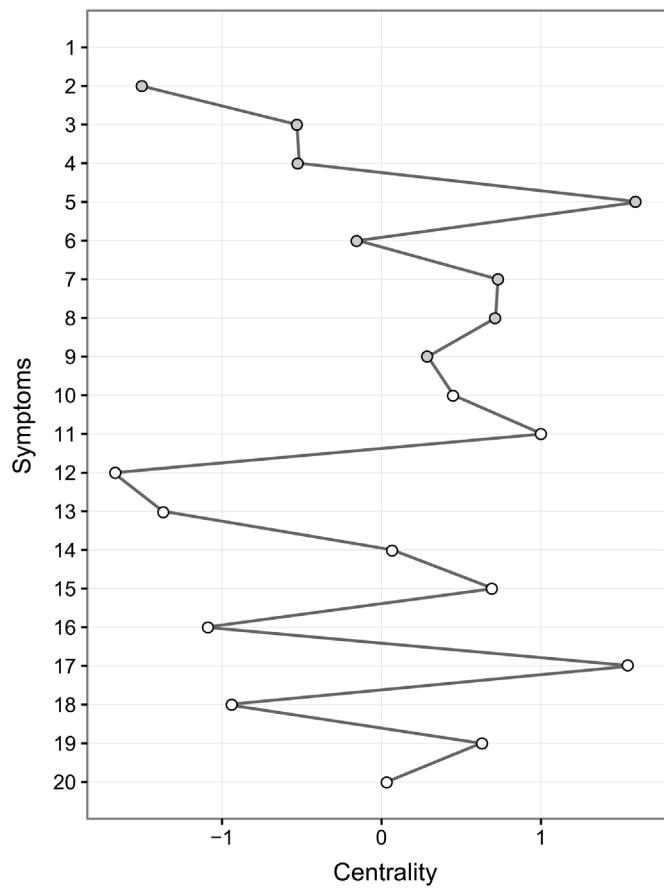
The major goal of this study was to examine, in a large, carefully assessed and ethnically homogeneous ascertained sample of severely depressed patients, the performance of the DSM criteria for MD compared to a selected set of non-DSM criteria judged by one of us (KSK) on the basis of clinical and research experience to be valuable in the evaluation of depressed patients. Using network analyses, we empirically investigated whether we could find support for or against the impressions gleaned from three historical (Kendler, 2016, 2017b; Kendler et al., 2010) and one empirical study (Fried, 2016) that the symptoms selected for inclusion in DSM are a sensible and clinically informed set of criteria but not necessarily unique or optimal in the assessment of depression.

The DSM criteria overall did a reasonable job reflecting the syndrome of MD as reported by these severely depressed Han Chinese women. However, consistent with these historical reviews, the network analyses indicated that the DSM criteria do not appear to be unique in characterizing depression when examined in the context of selected non-DSM depressive symptoms. Rather, the DSM and non-DSM symptoms appeared to be part of a single intermingled network with substantial inter-relations between them. We then performed a formal analysis comparing the centrality of the DSM and non-DSM symptoms in the network. The two sets of depressive symptoms did not differ statistically.

In two instances, we can link historical discussions about depressive symptoms with the results of our analyses. First, decreased sex drive was included as a depressive symptom in both the criteria proposed by Cassidy et al. (1957) and Feighner et al. (1972) but was not in the DSM-III definitions. Our analyses found that decreased libido was more central than the more traditional neurovegetative symptoms of changes



**Fig. 1.** Regularized partial correlation network of 9 DSM and 11 non-DSM depression symptoms. Connections between symptoms depict conservative estimates of partial correlations. Sad mood is not included in the network because it was endorsed by nearly everybody and thus showed too little variability.



**Fig. 2.** Standardized centrality estimates of the network of 9 DSM and 11 non-DSM depression symptoms. Centrality refers to the sum of all absolute connections of each symptom with all other symptoms. The centrality of sad mood is not available because it was not included in the network. Please refer to Fig. 1 for symptom shortcodes.

in weight/appetite and sleep. Second, anticipating later work by Beck et al. (1980), early clinicians described a wide variety of cognitive changes in depression including both hopelessness and feelings of inadequacy (Kendler, 2016, 2017b). Symptoms of hopelessness and low self-confidence were more broadly connected to other depressive

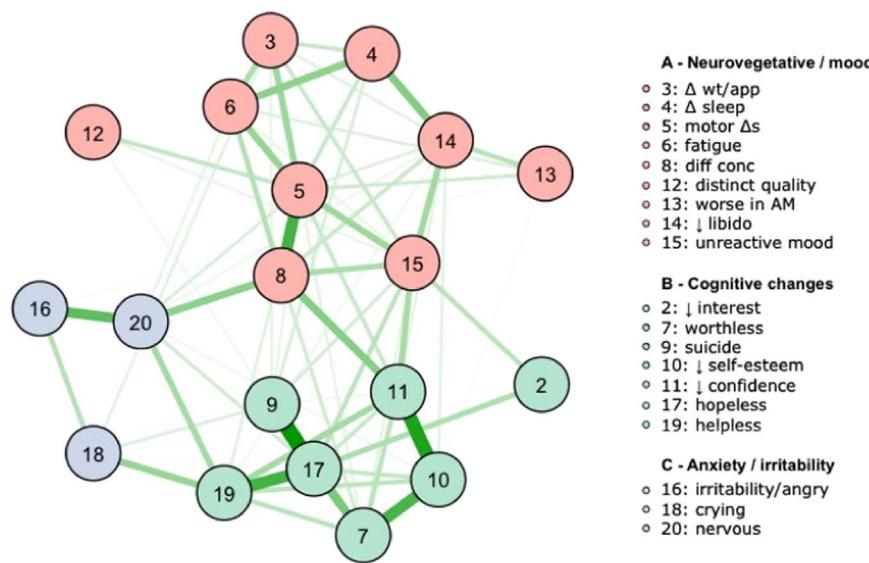
symptoms than the related DSM criterion of guilt/worthlessness.

Our findings can be further compared with the most similar prior study in the literature by two of us (EIF and DB) with colleagues (Fried et al., 2016). This study performed a network analysis similar to that applied here to 28 depressive symptoms assessed by the Inventory of Depressive Symptomatology (IDS) (Rush et al., 1996) in 3463 depressed outpatients meeting DSM-IV criteria for MD from the Sequenced Treatment Alternatives to Relieve Depression (STAR\*D) study (Fava et al., 2003). Of the 28 IDS items, 15 derived from DSM MD criteria and 13 did not. Because the symptom content for the non-DSM items in the IDS was quite different from the non-DSM items used in the CONVERGE study, a direct criterion-based comparison of these findings is not very useful. However, like us, they examined the centrality/inter-connectedness of the DSM and non-DSM items. Like us, they found the two did not differ significantly.

The most detailed prior study of the performance of the DSM-IV MD criteria was performed by Zimmerman et al. in a sample of 1800 psychiatric out-patients. Two results were particularly salient. First, among the DSM MD criteria, weight change was one of the two criteria that least frequently impacted on diagnosis, and thus could most easily be eliminated (Zimmerman et al., 2006). Second, among non-DSM criteria helplessness and hopelessness had better sensitivity and specificity at predicting MD case status than many of the DSM-IV MD criteria while anxiety was worse than the standard criteria (McGlinchey et al., 2006). Although the nature of the sample and the clinical severity differed substantially from the CONVERGE sample used here, some of these findings – especially with regard to the relatively poor, moderate and strong performances of, respectively, weight changes, anxiety symptoms and helplessness and hopelessness – replicated across samples.

Consistent with most prior studies, our findings suggest that the DSM criteria for MD are not unique or “more central” in their ability to identify the depressive syndrome. Rather, they appear to be relatively representative of a broader class of depression symptoms, some members of which might be similar to or even superior to individual DSM criteria in their capacity to capture critical features of MD.

We also examined the substructure of our network using two different community detection methods with comparable results. Our main finding – that most symptoms used to identify MD belong to two major subgroups – cognitive and neurovegetative – are consistent with several prior studies (Bringmann et al., 2015; Lux and Kendler, 2010; Wichers et al., 2005). In particular, a prior network analysis of the Beck Depression Inventory longitudinal scores from 182 depressed



**Fig. 3.** Results of the community detection analysis performed on the network in [Fig. 1](#). The three identified communities depict symptoms that more strongly inter-correlate with each other.

patients, a community structure analysis showed two groups of items identified as “cognitive” and “somatic-affective,” quite similar to our findings ([Bringmann et al., 2015](#)). Our detection of a third cluster of anxiety/irritability symptoms is consistent with the results of the prior historical review which found that 14/19 textbook authors described that symptoms of anxiety, apprehension or irritability as common clinical manifestations of the depressive syndrome ([Kendler, 2016](#)) as well as a range of clinical studies showing that anxiety and irritability are prevalent in depressed patients and associated both with poorer treatment response and chronicity ([Gollan et al., 2012](#); [Judd et al., 2013](#); [Van Loo et al., 2014](#)).

Our findings can be usefully compared with the results of a prior conventional exploratory and confirmatory factor analyses of an overlapping set of symptoms in the CONVERGE sample ([Li et al., 2014](#)). The factor analyses included disaggregated DSM criteria that separated increased from decreased appetite, weight and sleep. Two of the factors identified these typical versus atypical vegetative features. However, the remaining three factors had strong loadings on, respectively: i) helplessness, hopelessness, worthlessness, and suicidal ideation, ii) irritability, nervousness, and crying, and iii) psychomotor changes, difficulty concentrating, fatigue, and reduced libido. These results are reassuring in that different statistical methods with varying assumptions revealed substantially overlapping pictures of the symptomatic structure of severe depression.

Finally, one of us (KSK) has recently articulated two distinct ways in which to understand the relationship between DSM disorders and their diagnostic criteria: constitutive and indexical ([Kendler, 2017a](#)). In a constitutive relationship, criteria definitively define the disorder so that having a disorder is nothing more than meeting the criteria. In an indexical relationship, by contrast, the criteria are fallible indices of a disorder understood as a tentative or hypothetical diagnostic construct. Our current results provide further empirical support for the indexical position with respect to MD. DSM-III criteria are not unique or specific in the ways in which they reflect the depressive syndrome as would be required by the constitutive position. Rather, consistent with the indexical approach, our findings suggest that DSM criteria for MD are one subset of a broader number of possible criteria that could have been proposed by the relevant DSM panels.

#### 4.1. Limitations

These results should be interpreted in the context of three potentially important methodological limitations. First, our sample consisted of patients with relatively severe levels of illness and was restricted to

Han Chinese women. These results may not extrapolate to men, other ethnic groups, or to symptom relationships observed in the general population. However, the performance of the DSM-IV criteria for MD in this sample does not differ from comparable samples studies in North America or Europe ([Kendler et al., 2015](#)), thereby increasing the chances that our findings have broader applicability to depressed patients. Second, in our sample, selected for meeting DSM-IV MD criteria, the 8 DSM criteria analyzed had a significantly higher endorsement rate and therefore lower variance, than the non-DSM criteria (means: 0.92 and 0.82,  $t = 2.51$ ,  $df = 17$ ,  $p = 0.02$ ; SDs: 0.25 and 0.37,  $t = -2.79$ ,  $df = 12.18$ ,  $p = 0.02$ ). Could this bias our findings? We found that the correlation between the standard deviation and centrality of the analyzed symptoms was very low (+ 0.02), suggesting no consistent relationship that would bias the results. Third, of the analyzed symptoms, the DSM criteria “loss of interest” had the highest endorsement and lowest variance. This could have contributed to its peripheral placement in the network.

## 5. Conclusions

In a large sample of clinically depressed Han Chinese women, we performed a network analysis of DSM-IV criteria for MD along with a set of non-DSM depressive symptoms chosen for their clinical relevance. The resulting network has a structure that intermingled DSM and non-DSM symptoms. Furthermore, the “inter-connectedness” of the DSM criteria did not differ from the non-DSM symptoms. These results are consistent with the hypothesis, suggested by historical research, that the DSM criteria were chosen from a larger pool of possible items and while they perform well, they are not unique in their ability to capture the core nature of the depressive syndrome.

## Acknowledgments

This work was funded by the Wellcome Trust (WT090532/Z/09/Z, WT083573/Z/07/Z, WT089269/Z/09/Z) and by NIH grant MH100549. KSK and JF are part of the CONVERGE consortium (China, Oxford and Virginia Commonwealth University Experimental Research on Genetic Epidemiology) and gratefully acknowledge the support of all partners in hospitals across China. Special thanks to all the CONVERGE collaborators and patients who made this work possible. DB and EIF are supported by the European Research Council Consolidator grant no. 647209.

## Location of where work was done

Department of Psychological Methods, University of Amsterdam and Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University, Richmond, VA, USA.

## Ethical standards

The authors assert that all procedures contributing to this work comply with the ethical standards of the relevant national and institutional committees on human experimentation and with the Helsinki Declaration of 1975, as revised in 2008.

## Appendix A. Supplementary material

Supplementary data associated with this article can be found in the online version at <http://dx.doi.org/10.1016/j.jad.2017.11.032>.

## References

American Psychiatric Association, 1994. Diagnostic and Statistical Manual of Mental Disorders, Fourth ed. American Psychiatric Association, Washington, DC.

Beck, A.T., Rush, A.J., Shaw, B.F., Emery, G., 1980. Cognitive Therapy of Depression. John Wiley and Sons, London.

Bringmann, L.F., Lemmens, L.H., Huibers, M.J., Borsboom, D., Tuerlinckx, F., 2015. Revealing the dynamic network structure of the Beck Depression Inventory-II. *Psychol. Med.* 45, 747–757.

Cassidy, W.L., Flanagan, N.B., Spellman, M., Cohen, M.E., 1957. Clinical observations in manic-depressive disease: a quantitative study of one hundred manic-depressive patients and fifty medically sick controls. *J. Am. Med. Assoc.* 164, 1535–1546.

CONVERGE consortium, 2015. Sparse whole-genome sequencing identifies two loci for major depressive disorder. *Nature* 523, 588–591.

Epskamp, S., Borsboom, D., Fried, E.I., 2017. Estimating Psychological Networks and their Accuracy: A Tutorial Paper. *Behav. Res. Methods* 24 March[00], 1–34. Behavioral Research Methods[stat.AP] arXiv:1604.08462. 9–14–2016.

Epskamp, S., Cramer, A.O., Waldorp, L.J., Schmittmann, V.D., Borsboom, D., 2012. qgraph: network visualizations of relationships in psychometric data. *J. Stat. Softw.* 48, 1–18.

Epskamp, S., Fried, E.I., 2015. bootnet: Bootstrap methods for various network estimation routines. CRAN.r-project.org.

Fava, M., Rush, A.J., Trivedi, M.H., Nierenberg, A.A., Thase, M.E., Sackeim, H.A., Quitkin, F.M., Wisniewski, S., Lavori, P.W., Rosenbaum, J.F., Kupfer, D.J., 2003. Background and rationale for the sequenced treatment alternatives to relieve depression (STAR\*D) study. *Psychiatr. Clin. North Am.* 26, 457–494 (x).

Feighner, J.P., Robins, E., Guze, S.B., Woodruff Jr., R.A., Winokur, G., Munoz, R., 1972. Diagnostic criteria for use in psychiatric research. *Arch. Gen. Psychiatry* 26, 57–63.

Fried, E.I., 2016. The 52 symptoms of major depression: Lack of content overlap among seven common depression scales. *J. Affect. Disord.* 208, 191–197.

Fried, E.I., Epskamp, S., Nesse, R.M., Tuerlinckx, F., Borsboom, D., 2016. What are 'good' depression symptoms? Comparing the centrality of DSM and non-DSM symptoms of depression in a network analysis. *J. Affect. Disord.* 189, 314–320.

Fruchterman, T., Reingold, E., 1991. Graph drawing by force-directed placement. *Softw.: Pract. Exp.* 21, 1129–1164.

Gollan, J.K., Fava, M., Kurian, B., Wisniewski, S.R., Rush, A.J., Daly, E., Miyahara, S., Trivedi, M.H., 2012. What are the clinical implications of new onset or worsening anxiety during the first two weeks of SSRI treatment for depression? *Depress. Anxiety* 29, 94–101.

Judd, L.L., Schettler, P.J., Coryell, W., Akiskal, H.S., Fiedorowicz, J.G., 2013. Overt irritability/anger in unipolar major depressive episodes: past and current characteristics and implications for long-term course. *JAMA Psychiatry* 70, 1171–1180.

Kendler, K.S., 2016. The phenomenology of major depression and the Representativeness and nature of DSM criteria. *Am. J. Psychiatry* 173, 771–780.

Kendler, K.S., 2017a. DSM disorders and their criteria: how should they inter-relate? *Psychol. Med.* 1–7.

Kendler, K.S., 2017b. The genealogy of major depression: symptoms and signs of Melancholia from 1880–1900. *Mol. Psychiatry* 15 (00 [epub]).

Kendler, K.S., Aggen, S.H., Li, Y., Lewis, C.M., Breen, G., Boomsma, D.I., Bot, M., Penninx, B.W., Flint, J., 2015. The similarity of the structure of DSM-IV criteria for major depression in depressed women from China, the United States and Europe. *Psychol. Med.* 45, 1945–1954.

Kendler, K.S., Gardner, C.O., Neale, M.C., Prescott, C.A., 2001. Genetic risk factors for major depression in men and women: similar or different heritabilities and same or partly distinct genes? *Psychol. Med.* 31, 605–616.

Kendler, K.S., Munoz, R.A., Murphy, G., 2010. The development of the Feighner criteria: an historical perspective. *Am. J. Psychiatry* 167, 134–142.

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. The structure of the symptoms of major depression: exploratory and confirmatory factor analysis in depressed Han Chinese women. *Psychol. Med.* 44, 1391–1401.

Lux, V., Kendler, K.S., 2010. Deconstructing major depression: a validation study of the DSM-IV symptomatic criteria. *Psychol. Med.* 40, 1679–1690.

McGlinchey, J.B., Zimmerman, M., Young, D., Chelminski, I., 2006. Diagnosing major depressive disorder VIII: are some symptoms better than others? *J. Nerv. Ment. Dis.* 194, 785–790.

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

Pons, P., Latapy, M., 2005. Computing Communities in Large Networks using Random Walks (Long Version). Cornell University Library. Physics and Society.

Reichardt, J., Bornholdt, S., 2006. Statistical mechanics of community detection. *Phys. Rev. E. Stat. Nonlinear Soft. Matter Phys.* 74, 016110–1–016110–14.

Rush, A.J., Gullion, C.M., Basco, M.R., Jarrett, R.B., Trivedi, M.H., 1996. The Inventory of depressive symptomatology (IDS): psychometric properties. *Psychol. Med.* 26, 477–486.

Spitzer, R.L., Endicott, J., Robins, E., 1975. Research Diagnostic Criteria for a Selected Group of Functional Disorders. New York Psychiatric Institute, New York.

Stone, T.T., Burris, B.C., 1950. Melancholia; clinical study of 50 selected cases. *J. Am. Med. Assoc.* 142, 165–168.

Tibshirani, R., 1996. Regression shrinkage and selection via the Lasso. *J. R. Stat. Soc. Ser. B-Methodol.* 58, 267–288.

van Borkulo, C.D., Borsboom, D., Epskamp, S., Blanken, T.F., Boschloo, L., Schoevers, R.A., Waldorp, L.J., 2014. A new method for constructing networks from binary data. *Sci. Rep.* 4, 1–10.

van Borkulo, C.D., Epskamp, S., 2016. IsingFit: Fitting Ising Models Using the ELasso Method; Version 0.3.1. with contributions from Alexander Robitzsch. [IsingSampler]. CRAN.r-project.org. 11–2–2016.

Van Loo, H.M., Cai, T., Gruber, M.J., Li, J., de, J.P., Petukhova, M., Rose, S., Sampson, N.A., Schoevers, R.A., Wardenaar, K.J., Wilcox, M.A., Al-Hamzawi, A.O., Andrade, L.H., Bromet, E.J., Bunting, B., Fayyad, J., Floreescu, S.E., Gureje, O., Hu, C., Huang, Y., Levinson, D., Medina-Mora, M.E., Nakane, Y., Posada-Villa, J., Scott, K.M., Xavier, M., Zarkov, Z., Kessler, R.C., 2014. Major depressive disorder subtypes to predict long-term course. *Depress. Anxiety* 31, 765–777.

Wichers, M.C., Koek, G.H., Robaeys, G., Praamstra, A.J., Maes, M., 2005. Early increase in vegetative symptoms predicts IFN-alpha-induced cognitive-depressive changes. *Psychol. Med.* 35, 433–441.

World Health Organization, 1990. Composite International Diagnostic Interview (CIDI, Version 1.0). WHO, Geneva.

Zimmerman, M., McGlinchey, J.B., Young, D., Chelminski, I., 2006. Diagnosing major depressive disorder III: can some symptoms be eliminated from the diagnostic criteria? *J. Nerv. Ment. Dis.* 194, 313–317.