

Original Article

Cite this article: Lee CT, Kelley SW, Palacios J, Richards D, Gillan CM (2024). Estimating the prognostic value of cross-sectional network connectivity for treatment response in depression. *Psychological Medicine* **54**, 317–326. <https://doi.org/10.1017/S0033291723001368>

Received: 7 December 2022

Revised: 6 April 2023

Accepted: 25 April 2023

First published online: 7 June 2023


Keywords:

Anxiety; centrality; depression; internet-delivered cognitive behavioural therapy; network analysis; network connectivity; real-world; treatment prediction; treatment prognosis; treatment response

Corresponding author:

Chi Tak Lee; Email: Clee9@tcd.ie

Estimating the prognostic value of cross-sectional network connectivity for treatment response in depression

Chi Tak Lee^{1,2,3} , Sean W. Kelley^{1,2,3}, Jorge Palacios¹, Derek Richards^{1,4} and Claire M. Gillan^{1,2,3}

¹Department of Psychology, Trinity College Dublin, Dublin, Ireland; ²Trinity College Institute of Neuroscience, Trinity College Dublin, Dublin, Ireland; ³Global Brain Health Institute, Trinity College Dublin, Dublin, Ireland and ⁴SilverCloud Science, SilverCloud Health Ltd, Dublin, Ireland

Abstract

Background. Tightly connected symptom networks have previously been linked to treatment resistance, but most findings come from small-sample studies comparing single responder *v.* non-responder networks. We aimed to estimate the association between baseline network connectivity and treatment response in a large sample and benchmark its prognostic value against baseline symptom severity and variance.

Methods. $N = 40\,518$ patients receiving treatment for depression in routine care in England from 2015–2020 were analysed. Cross-sectional networks were constructed using the Patient Health Questionnaire-9 (PHQ-9) for responders and non-responders ($N = 20\,259$ each). To conduct parametric tests investigating the contribution of PHQ-9 sum score mean and variance to connectivity differences, networks were constructed for 160 independent subsamples of responders and non-responders (80 each, $n = 250$ per sample).

Results. The baseline non-responder network was more connected than responders (3.15 *v.* 2.70 , $S = 0.44$, $p < 0.001$), but effects were small, requiring $n = 750$ per group to have 85% power. Parametric analyses revealed baseline network connectivity, PHQ-9 sum score mean, and PHQ-9 sum score variance were correlated ($r = 0.20$ – 0.58 , all $p < 0.001$). Both PHQ-9 sum score mean ($\beta = -1.79$, $s.e. = 0.07$, $p < 0.001$), and PHQ-9 sum score variance ($\beta = -1.67$, $s.e. = 0.09$, $p < 0.001$) had larger effect sizes for predicting response than connectivity ($\beta = -1.35$, $s.e. = 0.12$, $p < 0.001$). The association between connectivity and response disappeared when PHQ-9 sum score variance was accounted for ($\beta = -0.28$, $s.e. = 0.19$, $p = 0.14$). We replicated these results in patients completing longer treatment (8–12 weeks, $N = 22\,952$) and using anxiety symptom networks ($N = 70\,620$).

Conclusions. The association between baseline network connectivity and treatment response may be largely due to differences in baseline score variance.

Introduction

Depression is one of the most prevalent and debilitating mental illnesses worldwide (World Health Organization, 2023). Despite the availability of effective treatments, a considerable proportion of individuals with depression still fail to achieve an adequate and sustained improvement (Cuijpers et al., 2021; Rush et al., 2006; Trivedi et al., 2006). Unfortunately, there are few reliable and robust characteristics that distinguish those who respond and do not respond to treatment with most efforts to-date focused on standardised clinical measures and demographics (Maj et al., 2020; McMahon, 2014; Rost, Binder, & Brückl, 2023). Many have suggested this is a consequence of the way we conceptualise depression as a latent phenomenon that causes observed symptoms like sadness and anhedonia, which we typically sum to produce an overall depression score. Network theory of psychopathology forwards a different perspective and posits that symptoms are interacting components of a dynamical system (Borsboom, 2017; Borsboom & Cramer, 2013), which can result in positive feedback loops that propel people into episodes of illness. The greater the connectivity of these symptom networks, the lower the psychological resilience one has, with more connected networks reacting more strongly to perturbations and taking longer to recover. A key prediction of network theory emerging from this is that individuals with tightly connected networks should have greater vulnerability to depression, poorer prognosis, and more treatment resistance (Cramer et al., 2016; Pe et al., 2015; van Borkulo et al., 2015).

Several studies tested this using cross-sectional network analysis. van Borkulo and colleagues compared baseline connectivity differences between persisters ($n = 253$) and remitters ($n = 262$) of depression after two years (van Borkulo et al., 2015). In line with network theory, persisters had tighter network connectivity compared to remitters at baseline. This was

© The Author(s), 2023. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

replicated in a child and adolescent sample ($n = 566/174$) (McElroy, Napoleone, Wolpert, & Patalay, 2019), but there have also been null findings, for example, in adolescents ($n = 232/233$) (Schworen, van Borkulo, Fried, & Goodyer, 2018), and when depression and anxiety symptoms were examined together ($n = 956/1466$) (O'Driscoll *et al.*, 2021). On a more granular level, some studies have shown that the severity of symptoms that are more 'central' (i.e. important) is associated with non-response (Elliott, Jones, & Schmidt, 2020; Hagan *et al.*, 2021), and that improvements in central symptoms predict changes in other symptoms (Papini, Rubin, Telch, Smits, & Hien, 2020; Robinaugh, Millner, & McNally, 2016; Rodebaugh *et al.*, 2018). Findings regarding the centrality hypothesis, however, are not univocal (O'Driscoll *et al.*, 2021; Spiller *et al.*, 2020), and it remains unclear whether centrality measures perform better than other network/non-network metrics when compared directly. Finally, contrary to network theory, a host of studies have reported that connectivity increases (rather than decreases) after treatment (Beard *et al.*, 2016; Berlim, Richard-Devantoy, Dos Santos, & Turecki, 2021; Blanco *et al.*, 2020; Bos *et al.*, 2018; Curtiss *et al.*, 2021).

One of the common critiques of the network literature is the over-reliance on cross-sectional data and methods; estimating correlations between symptoms across-subjects rather than within-subject (Contreras, Nieto, Valiente, Espinosa, & Vazquez, 2019; Robinaugh, Hoekstra, Toner, & Borsboom, 2020), and often employing small samples (Schumacher, Burger, Echterhoff, & Kriston, 2022). This introduces two issues. First, it is uncertain if cross-sectional relationships between symptoms correspond to intraindividual relationships (Epskamp & Fried, 2018; Fisher, Medaglia, & Jeronimus, 2018). Second, cross-sectional studies typically construct just two networks of differential treatment response for comparison. This precludes controlling for potential confounds such as symptom severity and variance. Variance is of particular interest as it relates to the strength of the association that can be observed between symptoms. Cross-sectional networks are typically estimated from the partial correlations between symptom-pairs (Fried *et al.*, 2016), and the correlation between any two symptoms is their covariance proportional to their total variance. This leaves network estimation susceptible to differences in variance, which can be introduced artificially when creating sub-groups of participants (Bos & De Jonge, 2014; Fried *et al.*, 2016; Terluin, de Boer, & de Vet, 2016). Prior research has shown that connectivity differences remain when groups are matched on baseline severity (McElroy *et al.*, 2019; van Borkulo *et al.*, 2015), but to our knowledge, none have assessed the impact of differences in variance.

Our study sought to fill this gap by examining baseline network differences in $N = 40\,518$ patients who received internet-delivered cognitive behavioural therapy (iCBT) for depression. Leveraging our large sample, we adopted a novel subsampling approach so to conduct parametric analyses for 160 independent responder and non-responder networks with $n = 250$ unique patients per subsample. Importantly, these subsamples naturally varied in levels of baseline network connectivity, symptom severity, and variance. This allowed us to assess if differences in cross-sectional network connectivity are better explained by differences in depression severity and/or variance, which have not been separable using standard methods comparing single dyads of responder-non-responder networks. Additionally, using the independent networks from the subsampling method, we assessed whether other network metrics such as symptom centrality related

to treatment success and contextualised their effect sizes against simpler metrics such as mean and variance of individual symptoms. Finally, findings were tested for generalisation to partially overlapping samples receiving iCBT of a longer duration (8–12 weeks) and to networks constructed from anxiety symptoms in patients receiving anxiety-relevant iCBT.

Methods

Study setting and intervention

We examined an archival dataset of patients who received iCBT from SilverCloud Health between January 2015 and December 2020, as part of the Improving Access to Psychological Therapies programme within the National Health Service in England. The intervention followed NICE guidelines and have shown efficacy in improving clinical outcomes with sustained effects (Palacios *et al.*, 2022; Richards *et al.*, 2020). Patients provided their consent for their anonymised data to be used in routine service evaluations.

Outcome measure

Depression was measured by the Patient Health Questionnaire-9 (PHQ-9) (Kroenke, Spitzer, & Williams, 2001). PHQ-9 was administered to patients at the beginning of each iCBT session, but patients were able to skip and return to these assessments later.

Study sample

Figure 1a illustrates the process from which we derived our final study sample. First, patients were excluded if they did not have at least one PHQ-9 completed in a timeframe of 4–8 weeks since treatment initiation. We included patients completing relatively short durations of treatment (i.e. 4 weeks) due to the self-paced nature of iCBT (Lawler, Earley, Timulak, Enrique, & Richards, 2021). The last PHQ-9 completed within the 4–8 week window was deemed as the follow-up assessment. As the study focused on examining the association between depression network characteristics and clinical changes following treatment for depression, patients were further excluded if they scored <10 on the PHQ-9 (i.e. did not reach 'caseness' for depression) at baseline, and if they were enrolled in any other type of iCBT programme not purposed for treating depression. Most patients were clinician-guided, meaning treatment progress was monitored and facilitated by a clinician. As prior studies have shown differential efficacy of iCBT when guided *v.* unguided (Karyotaki *et al.*, 2021), we excluded data from patients who were unguided. Furthermore, patients who satisfied the responder and non-responder status defined in our study were included. Patients were classified as Responder if (1) they recovered (i.e. transitioned from 'caseness' to 'non-caseness' post-treatment), and (2) their score reduction was greater than the Reliable Change Index of ≥ 6 on the PHQ-9 (Jacobson & Truax, 1991). Patients were classified as Non-Responders if they met neither of these criteria, and patients who met only one of these criteria were treated as intermediate cases that were removed from analyses. Finally, as network estimation is influenced by sample size (Burger *et al.*, 2022), we yielded equal-sized groups for Responders and Non-Responders by matching the cohorts using 1:1 propensity score matching ($n = 20\,259$ per group), where each patient with a specific number of days in treatment in the Responder group was matched to

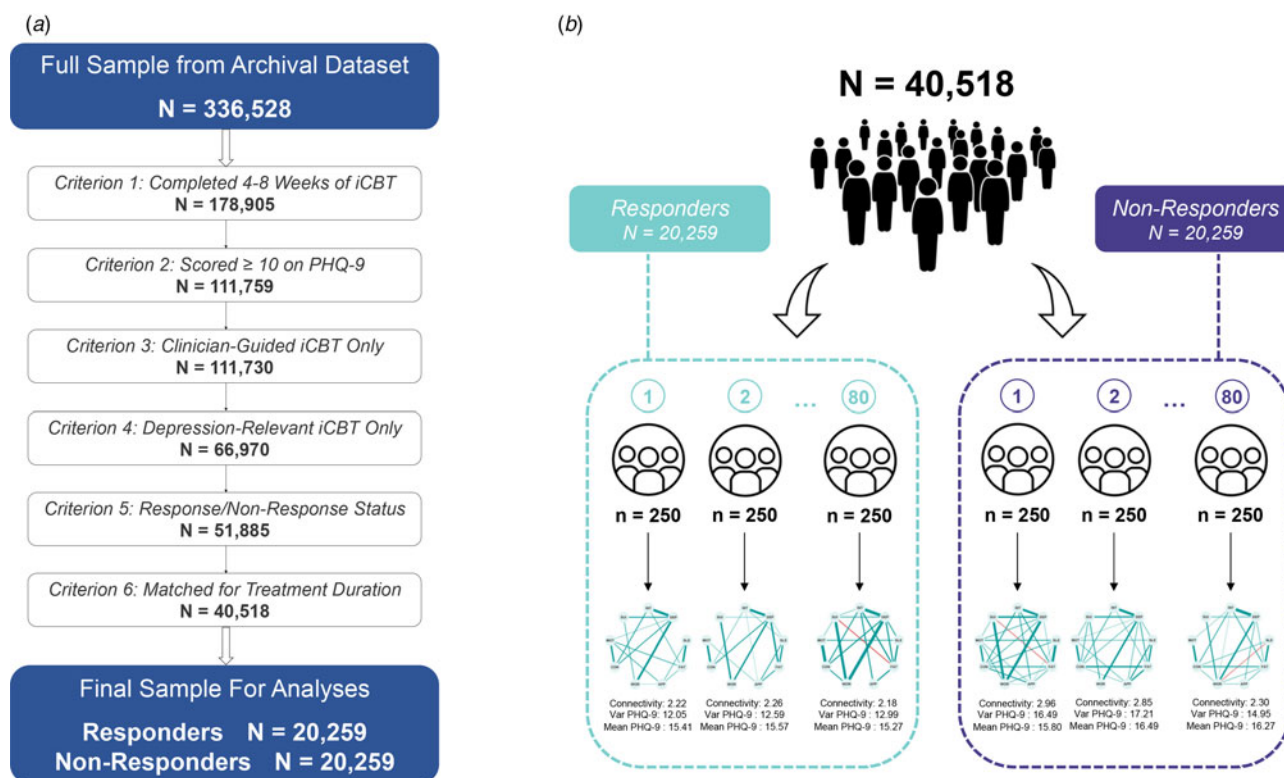


Figure 1. Sampling procedures for analyses. (a) Final study sample flow chart with inclusion and exclusion criteria. (b) Subsampling procedure for parametric analyses testing whether baseline depression severity and variance explained the association between network connectivity and treatment response. The Responder and Non-Responder samples were divided into 80 sets of $n = 250$, respectively, where each set differed naturally in PHQ-9 baseline mean and variance. Note: INT, 'loss of interest/pleasure'; DEP, 'depressed mood'; SLE, 'sleep'; FAT, 'fatigue'; APP, 'appetite'; WOR, 'worthlessness'; CON, 'concentration'; MOT, 'psychomotor problems'; SUI, 'suicidality'.

another patient with the same number of treatment days in the Non-Responder group, independent of their clinical scores.

Data analysis

Baseline and pre-post score analyses

Differences in PHQ-9 sum and item scores at baseline and follow-up, along with treatment engagement, were compared across Responders and Non-Responders using t tests and ANOVA.

Network analysis

Cross-sectional networks using Gaussian Graphical Models were estimated for Responders and Non-Responders at baseline and follow-up using all items of the PHQ-9 (Epskamp & Fried, 2018; Epskamp, Waldorp, Möttus, & Borsboom, 2018). Relationships between symptoms (nodes) were estimated using partial correlations (edges) (i.e. the relationship between two symptoms after controlling for the others within the same network). The glasso regularisation penalisation technique based on the Extended Bayesian Information Criterion was performed during model selection (Chen & Chen, 2008). A tuning hyperparameter ($\gamma = 0.5$) was employed to find the optimal balance between parsimony and goodness of fit of the network. Network connectivity was defined as the weighted sum of the signed associations between nodes. For symptom centrality, we focused on examining node strength as one of the most evaluated and intuitive metrics in psychological networks. It quantifies the

strength of a node's direct connections to other nodes in the network (Bringmann, 2021). Statistical significance testing on network connectivity, edge-specific, and centrality differences were conducted using the Network Comparison Test (NCT) (van Borkulo et al., 2022). The NCT is a two-tailed resampling-based permutation test that compares network differences between two independent cross-sectional networks (responders and non-responders). Edge-difference networks (i.e. subtracting two network covariance-matrices) were used to illustrate significant edge differences between networks.

Power estimation

To determine the required sample size to detect connectivity differences between Responders *v.* Non-Responders at baseline, we repeated the NCT 1000 times for random subsets of $n = 250$, $n = 500$, $n = 750$, and $n = 1000$ per group and reported the statistical power, i.e., the proportion of samples in which a significant difference was detected.

Subsampling analysis

To test if the relationship between connectivity and treatment response is explained by baseline severity and/or variance differences, we divided our sample into 160 independent subsamples of $n = 250$, of which 80 were Responders and 80 were Non-Responders (Fig. 1b). Each subsample naturally varied in baseline PHQ-9 sum score mean and PHQ-9 sum score variance, which allowed us to treat these networks as unique observations in linear regressions predicting network connectivity from

response status, with baseline PHQ-9 sum score mean and PHQ-9 sum score variance as covariates. Using these independent subsamples, we further tested the added prognostic value of network metrics for treatment success; we contextualised the magnitude (i.e. effect size) of the association between baseline network connectivity and treatment response by comparing it to other baseline measures in univariate regressions, with response status as the IV, and the severity and variance of PHQ-9 sum and item score as well as strength centrality of individual symptoms as DVs. We repeated this procedure to test for differences in network connectivity, prior to and after treatment.

Generalisation test

To test if our main results generalised, we applied the same analytical procedures to two other samples from our dataset. This included (1) a smaller group of patients ($N = 22\,952$) who underwent a longer course of iCBT (8–12 weeks) for depression to examine treatment duration effect (online Supplementary eAppendix 1) and (2) a larger group of patients ($N = 70\,620$) who received iCBT for anxiety to probe whether observed findings were disorder-specific, where response status and networks were based on the Generalised Anxiety Disorder-7 (GAD-7) (Spitzer, Kroenke, Williams, & Löwe, 2006) (online Supplementary eAppendix 2). The main dataset partially overlapped with both of these datasets (33% for the 8–12 weeks iCBT sample, 49% for the GAD sample).

All data processing and analyses were conducted using *R* (version 4.1.1). We used specific packages such as *MatchIt* for group matching (Ho, Imai, King, & Stuart, 2011), *qgraph* for network visualisation (Epskamp, Cramer, Waldorp, Schmittmann, & Borsboom, 2012), *bootnet* for network estimation (Epskamp & Fried, 2018), and *NetworkComparisonTest* for network comparisons (van Borkulo et al., 2022).

Results

Sample characteristics

Non-Responders had significantly higher baseline PHQ-9 sum score mean and PHQ-9 sum score variance ($M = 16.26$, $S.D. = 4.03$) compared to Responders ($M = 15.33$, $S.D. = 3.56$) [mean difference: $t(40\,516) = 24.64$, $p < 0.001$; variance difference: $F = 1.28$, $p < 0.001$]. Non-Responders also scored higher on all PHQ-9 items and had greater variance in ‘loss of interest/pleasure’, ‘depressed mood’, ‘psychomotor problems’, and ‘suicidality’ (Table 1, online Supplementary eFig. 1). By definition, Responders exhibited a larger reduction post-treatment ($M = 10.06$, $S.D. = 3.47$) than Non-Responders ($M = 0.13$, $S.D. = 3.36$) in PHQ-9 sum score, $t(40\,516) = 292.83$, $p < 0.001$, even after controlling for imbalance in baseline PHQ-9 sum score mean, $F_{(1, 40515)} = 121\,473.12$, $p < 0.001$ (Fig. 2a, online Supplementary eTable 1, eFig. 2). On average, Responders were in treatment one day longer ($M = 44.17$, $S.D. = 7.93$) than Non-Responders ($M = 43.07$, $S.D. = 8.22$), $t(51\,883) = -15.36$, $p < 0.001$. There were more Non-Responders (68%) receiving depression-only iCBT *v.* comorbid depression-anxiety iCBT than Responders (65%), $\chi^2 = 64.09$ (2), $p < 0.001$.

Full sample network differences at baseline

The Non-Responder network had greater connectivity than the Responder network at baseline (3.15 *v.* 2.70 , $S = 0.44$, $p < 0.001$)

(Fig. 2b). This effect was small; a power analysis revealed that $n = 750$ per group was required to achieve 85% power to detect this (Fig. 2c). When we further matched both groups on baseline PHQ-9 sum scores, thereby matching on both PHQ-9 sum score mean and PHQ-9 sum score variance [$n = 18\,281$ per group; mean difference: $t(36\,560) = 0$, $p = 1$; variance difference: $F = 1.00$, $p = 1$], connectivity differences disappeared between Responders and Non-Responders (2.73 *v.* 2.72 , $S = 0.008$, $p = 0.80$), suggesting that sum score mean and/or variance drive the effect. We found 10/36 edges were significantly different between-groups (all $p < 0.05$) (online Supplementary eTable 3, eFig. 4a). The Non-Responder network had two more edges present, while the Responder network had five weaker positive edges and two stronger negative edges. With regards to strength centrality (Fig. 2d), ‘depressed mood’ was the most central symptom in both networks (1.18 *v.* 1.22 , $p = 0.17$). Responders exhibited greater strength in ‘worthlessness’ (0.93 *v.* 0.83 , $p = 0.004$) and ‘loss of interest/pleasure’ (0.89 *v.* 0.84 , $p = 0.047$), while ‘sleep’ (0.66 *v.* 0.74 , $p = 0.02$) and ‘psychomotor problems’ (0.61 *v.* 0.71 , $p = 0.002$) were significantly more central in the Non-Responder network (online Supplementary eTable 2).

Parametric analysis of PHQ-9 sum score mean, variance and network connectivity

Responders and Non-Responders differed in baseline PHQ-9 sum score and symptom mean, PHQ-9 sum score and symptom variance, and network connectivity. To disentangle these features, we drew 160 independent samples of $n = 250$ Responders and $n = 250$ Non-Responders (i.e. 80 subsets per group). We found that baseline PHQ-9 sum score mean and PHQ-9 sum score variance were positively correlated in the networks of both Responders ($r = 0.47$, $p < 0.001$) and Non-Responders ($r = 0.35$, $p = .002$), where the greater the PHQ-9 sum score means within each subsample, the higher the PHQ-9 sum score variances (Fig. 3a). We estimated networks for each subsample and found that networks were more connected in Non-Responders ($\beta = -1.35$, $S.E. = 0.11$, $p < 0.001$) (Fig. 3d). However, network connectivity across these subsamples was positively associated with baseline PHQ-9 sum score mean (Non-Responders, $r = 0.23$, $p = 0.04$; Responders, $r = 0.20$, $p = 0.08$; Fig. 3b) and PHQ-9 sum score variance (Responders $r = 0.41$; $p < 0.001$; Non-Responders $r = 0.58$, $p < 0.001$; Fig. 3c). Taking these network characteristics forward to a multiple linear regression analysis, group differences in network connectivity survived after controlling for baseline PHQ-9 sum score mean ($\beta = -0.71$, $S.E. = 0.26$, $p = 0.007$, Fig. 3e), but not PHQ-9 sum score variance ($\beta = -0.28$, $S.E. = 0.19$, $p = 0.14$, Fig. 3f).

Parametric analysis of symptom-level data

The subsampling analysis further revealed between-group differences in symptom strength, where the centrality of all symptoms were higher in the Non-Responder subsets (all $p < 0.001$, Fig. 3g). To contextualise these differences, we compared their effect sizes relative to the mean and variance of individual symptoms, and the aggregate measures from the prior section. We found that baseline PHQ-9 sum score mean was the most strongly associated with response status ($\beta = -1.79$, $S.E. = 0.07$, $p < 0.001$), with Non-Responders having greater baseline severity. This was followed by ‘suicidality’ mean ($\beta = -1.74$, $S.E. = 0.08$, $p < 0.001$), and baseline PHQ-9 sum score variance ($\beta = -1.67$, $S.E. = 0.09$,

Table 1. Comparisons of PHQ-9 item and sum score means and variances of Responders and Non-Responders at baseline

PHQ-9 Variable	Response Status		<i>t</i> / <i>F</i> (<i>p</i>)
	Responders (<i>N</i> = 20 259)	Non-Responders (<i>N</i> = 20 259)	
Loss of Interest/Pleasure			
Mean	1.90	1.96	8.27 (<i>p</i> < 0.001)
Variance	0.62	0.67	1.07 (<i>p</i> < 0.001)
Depressed Mood			
Mean	2.01	2.06	7.28 (<i>p</i> < 0.001)
Variance	0.60	0.63	1.06 (<i>p</i> < 0.001)
Sleep			
Mean	2.18	2.29	13.28 (<i>p</i> < 0.001)
Variance	0.75	0.70	0.93 (<i>p</i> < 0.001)
Fatigue			
Mean	2.32	2.42	13.69 (<i>p</i> < 0.001)
Variance	0.56	0.52	0.93 (<i>p</i> < 0.001)
Appetite			
Mean	1.76	1.90	14.03 (<i>p</i> < 0.001)
Variance	0.98	0.92	0.95 (<i>p</i> < 0.001)
Worthlessness			
Mean	2.02	2.07	6.07 (<i>p</i> < 0.001)
Variance	0.75	0.74	0.99 (<i>p</i> = 0.35)
Concentration			
Mean	1.84	1.72	12.92 (<i>p</i> < 0.001)
Variance	0.80	0.80	1.00 (<i>p</i> = 0.89)
Psychomotor Problems			
Mean	0.92	1.05	14.40 (<i>p</i> < 0.001)
Variance	0.85	0.93	1.10 (<i>p</i> < 0.001)
Suicidality			
Mean	0.51	0.66	20.01 (<i>p</i> < 0.001)
Variance	0.54	0.72	1.35 (<i>p</i> < 0.001)
PHQ-9 Total			
Mean	15.33	16.26	24.64 (<i>p</i> < 0.001)
Variance	12.70	16.21	1.28 (<i>p</i> < 0.001)

Note: All *p* values indicated above for PHQ-9 item comparisons have been adjusted for multiple significance testing using the Hochberg method.

p < 0.001) (Fig. 3g). Notably, the mean score of every symptom (except 'depressed mood') was more associated with treatment response than its centrality. The strength of 'depressed mood', the most central symptom at baseline for both groups, had the highest signal for treatment response of all other symptom strengths, but was still weaker than 7/9 measures of item means.

Network connectivity changes following treatment

Examining changes following treatment, the overall network connectivity of the full sample increased from baseline to follow-up (2.97 *v.* 4.08, *S* = 1.10, *p* < 0.001). These effects were evident separately in both the Responder networks (2.70 *v.* 3.25, *S* = 0.55,

p < 0.001) (online Supplementary eFigs 3a, 3c, 4c; eTable 4), and Non-Responder networks (3.15 *v.* 3.52, *S* = 0.38, *p* < 0.001) (online Supplementary eFigs 3b, 3d, 4d, eTable 5). At follow-up, Non-Responders continued to have a more connected network (3.52 *v.* 3.25 *S* = 0.27, *p* < 0.001) compared to Responders (online Supplementary eFigs 3c, 3d, 4b, eTable 6). In the subsampling analysis[†], we examined network connectivity in both groups, pre- and post-treatment. A repeated measures ANOVA revealed a significant main effect of Group, where Non-Responders had overall more connected networks, $F_{(1, 156)} = 197.23$, *p* < 0.001. There was also an effect of Time, where networks increased

[†]The notes appear after the main text.

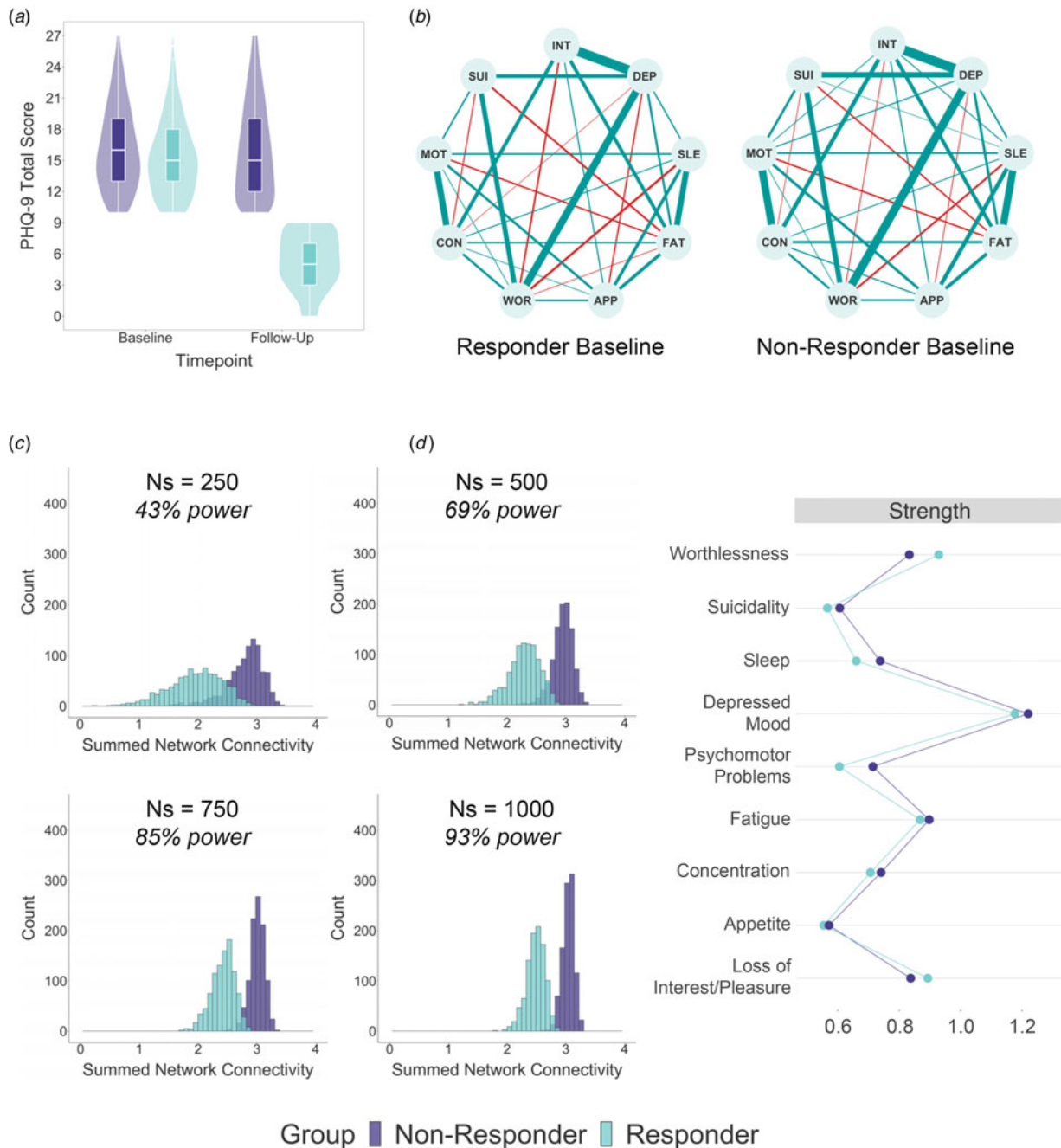


Figure 2. Full-Sample Network Differences at Baseline. (a) PHQ-9 sum score before and after iCBT by responder group. (b) Baseline network visualisations by responder group, where green and red edges denote positive and negative partial correlations, respectively. (c) Power analyses for detecting network connectivity differences at baseline. Responder and Non-Responder groups were randomly subsampled 1000 times each at $N = 250$, $N = 500$, $N = 750$, and $N = 1000$. (d) Strength centrality comparisons of PHQ-9 symptom nodes between Responders and Non-Responders at baseline.

Note: INT, 'loss of interest/pleasure'; DEP, 'depressed mood'; SLE, 'sleep'; FAT, 'fatigue'; APP, 'appetite'; WOR, 'worthlessness'; CON, 'concentration'; MOT, 'psychomotor problems'; SUI, 'suicidality'.

in connectivity from baseline to follow-up, $F_{(1, 156)} = 545.45$, $p < 0.001$. Finally, there was a Group by Time interaction, $F_{(1, 156)} = 37.44$, $p < 0.001$, driven by the fact that Responders had greater increases in connectivity ($M = 1.05$, $s.d. = 0.52$), $t(78) = -18.14$, $p.adj < 0.001$, compared to Non-Responders ($M = 0.62$, $s.d. = 0.37$), $t(78) = -14.79$, $p.adj < 0.001$. PHQ-9 sum score variance decreased over time for Responders but increased for Non-Responders, both likely a function of the small range of

values required to qualify for 'response' (online Supplementary eFig. 5). Correlational analyses revealed that changes in network connectivity were not associated with changes in PHQ-9 sum score mean for Responders ($r = 0.09$, $p = 0.44$) nor for Non-Responders ($r = 0.06$, $p = 0.60$). For both cohorts, changes in network connectivity were positively associated with changes in PHQ-9 sum score variance (Responder $r = 0.42$, $p < 0.001$; Non-Responder $r = 0.49$, $p < 0.001$).

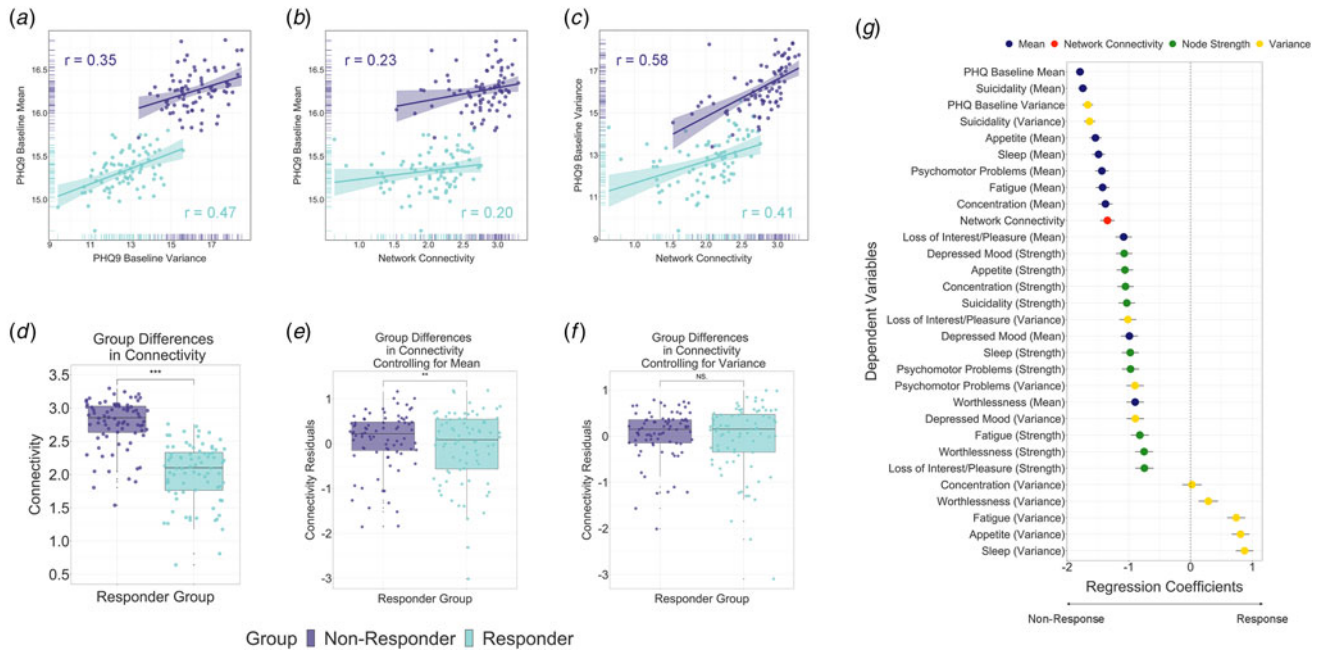


Figure 3. Parametric analyses on independent subsamples of Responders and Non-Responders. (a) Correlation between baseline PHQ-9 mean and variance of 160 independent samples of $N = 250$ participants used to construct networks, split by responder group. (b) The same analysis was carried out for baseline PHQ-9 mean and baseline network connectivity of these networks, and (c) baseline PHQ-9 variance and network connectivity. (d) Network connectivity differences between 80 Responder and 80 Non-Responders networks overall, and (e) after controlling for baseline PHQ-9 mean ($p = 0.007$) and (f) variance ($p = 0.14$), respectively. (g) Regression analyses with response status (Responder, Non-Responder) as IV and individual symptom features (mean, variance, centrality) as DVs. All regressions were statistically significant (all $p < 0.05$), except for variance in concentration ($p = 0.89$) and worthlessness ($p = 0.07$).

Replication and generalisation

To test the robustness of our main findings, we repeated the core analyses for two partially overlapping datasets including (1) patients receiving iCBT for 8–12 weeks ($N = 22\ 952$) and (2) where networks were based on anxiety symptoms ($N = 70\ 620$). We replicated our results in both sensitivity analyses: at baseline, the full sample Non-Responder network was more connected than the full sample Responder network in both patients undergoing longer treatment (3.08 v. 2.74 , $S = 0.34$, $p < 0.001$) and in those receiving iCBT for anxiety (2.68 v. 2.42 , $S = 0.26$, $p < 0.001$). Parametric analyses revealed that, in both cases, connectivity differences between Responders and Non-Responders were no longer significant when sum score variance was accounted for in the model (patients undergoing longer treatment: $\beta = 0.23$, $s.e. = 0.20$, $p = 0.25$; patients undergoing treatment for anxiety: $\beta = -0.19$, $s.e. = 0.12$, $p = 0.13$). Baseline sum score mean and sum score variance were also once again more predictive of treatment response than baseline network connectivity in both patients undergoing longer iCBT (mean: $\beta = -1.53$, $s.e. = 0.14$, $p < 0.001$; variance: $\beta = -1.50$, $s.e. = 0.14$, $p < 0.001$, connectivity: $\beta = -1.07$, $s.e. = 0.18$, $p < 0.001$) and those receiving iCBT for anxiety (mean: $\beta = -1.86$, $s.e. = 0.05$, $p < 0.001$; variance: $\beta = -1.78$, $s.e. = 0.06$, $p < 0.001$; connectivity: $\beta = -1.67$, $s.e. = 0.08$, $p < 0.001$). Lastly, correlational analyses examining network changes following treatment confirmed an association between sum score variance changes and network connectivity changes in both patients undergoing longer iCBT (Non-Responders: $r = 0.34$, $p = 0.02$; Responders: $r = 0.23$, $p = 0.127$) and patients receiving iCBT for anxiety (Non-Responder: $r = 0.59$, $p < 0.001$; Responder: $r = 0.49$, $p < 0.001$) (see online Supplementary eAppendix 1 and 2 for a detailed report of sensitivity analyses findings).

Discussion

Prior work has suggested that patients with more tightly connected symptom networks are more treatment resistant (Cramer et al., 2016; Pe et al., 2015; van Borkulo et al., 2015). However, existing studies are based on comparisons of single responder v. non-responder cross-sectional networks (Fisher et al., 2018), with relatively small samples (Forbes, Wright, Markon, & Krueger, 2017), that do not account for symptom variance (Terluin et al., 2016). We addressed these gaps in a sample of $N = 40\ 518$ that was analysed as a whole, and also divided into subsamples, thereby permitting parametric analyses of the role of variance in connectivity estimates, separate to that of severity. In the single network comparison, we found that connectivity was greater for Non-Responders than Responders at baseline. This effect was small, however, requiring $n = 750$ per group to reliably detect it, and we identified two potential confounds: Non-Responders had greater depression severity and variance at baseline. To disentangle these effects, we created 160 independent networks of Responders and Non-Responders ($n = 250$ each), and tested across networks if severity and/or variance explained connectivity differences. While the Non-Responder networks were on average more connected than the Responder networks at baseline, after controlling for sum score variance, the association between connectivity and treatment response was no longer significant. We replicated this result in two partially overlapping generalisation samples, one with patients undergoing a longer duration of iCBT (8–12 weeks) and another based on anxiety not depression (4–8 weeks).

This paper highlights an important confound that is understudied in the network literature (Bos & De Jonge, 2014; Terluin et al., 2016); network estimation is based on (partial)

symptom correlations which depend on the variance of these symptoms, not just their covariance. Imbalances in variance may be an inherent clinical characteristic of treatment-resistant groups (Friedman *et al.*, 2012), but can also be easily introduced when subgrouping patients based on treatment response (leading to range restriction of items) (Linn, 1968). That said, it is important to recognise that variance contributes to, but is not the same as, network connectivity. For example, correlations between network connectivity and variance were moderate ($r = 0.41-0.58$), and recent work examining temporal, intraindividual networks found associations with symptom change over time that survived controlling for variance (Kelley *et al.*, 2022). Moreover, networks actually became more connected following treatment despite reductions in both symptom severity and variance, a counter-intuitive finding most consistent in the network literature (Beard *et al.*, 2016; Berlim *et al.*, 2021; Blanco *et al.*, 2020; Bos *et al.*, 2018; Curtiss *et al.*, 2021). One explanation is that increased symptom connectivity is not necessarily bad; a highly connected network should theoretically lead to a more malleable system, but not necessarily worsening mental health (Fried *et al.*, 2016; McElroy *et al.*, 2019), as therapeutic gains may be due to systems becoming less 'stuck' and more open to change. Recent work examining personalised network dynamics in healthy individuals supports this; those with more connected depression networks tended to have greater fluctuations in depression over 6 weeks, but these went in both positive and negative directions of change (Kelley *et al.*, 2022).

Consistent with prior work (Hagan *et al.*, 2021; O'Driscoll *et al.*, 2021; Robinaugh *et al.*, 2016), 'depressed mood' had the highest strength centrality for both groups at baseline. However, we found that the severity of all symptoms (except 'worthlessness') was more strongly linked to treatment response than the strength of 'depressed mood', and that the severity of each symptom was more predictive than its corresponding strength centrality. Both baseline severity and variance were also more predictive of treatment success than network connectivity. The lack of added prognostic value of network metrics was previously challenged by Spiller *et al.* (2020), who found that both mean symptom severity and count were more predictive of symptom changes than centrality indices. Together these findings question the real-world prognostic utility of cross-sectional network metrics, over and beyond basic self-report symptom data readily available at baseline.

There are several limitations of this study. Firstly, this was a retrospective, observational study with no control group. Information on patient demographics and concurrent treatment such as medication status were also not available. It therefore remains unclear whether the observed results can be generalised to networks estimated with patients undergoing alternative treatment (e.g. antidepressant medication). Our main study sample was also limited to patients who, on average, scored on the cusp of the cut-off for determining less/more severe depression at baseline (i.e. 16 on the PHQ-9), and therefore may not be representative of all patients with depression enrolled in primary care (NICE, 2022). In addition, while the PHQ-9 is widely used for detecting and monitoring depression symptoms within routine care settings, the instrument is primarily purposed for screening depression symptoms against the Diagnostic and Statistical Manual of Mental Disorders-Fourth Edition (DSM-IV) by combining related symptomology into single items (Harrison *et al.*, 2021). Future research should consider gold-standard symptom assessments such as semi-structured clinical interviews

specifically designed and validated for in-depth assessment of individual symptoms of depression (e.g. Wing *et al.*, 1990). As previously noted, networks estimated from cross-sectional data do not always generalise onto an individual-level (Hamaker, 2012; von Klipstein, Borsboom, & Arntz, 2021), and indeed, differences in baseline sum score mean and sum score variance can be introduced systematically by the binary definition of 'response' that is required for cross-sectional network analysis. The crucial next step for network theory is to move towards a dynamical account of psychopathology afforded by personalised, within-subject networks for each patient undergoing treatment overtime.

Conclusion

In a large sample of > 40 000 patients, we determined that network connectivity differences between iCBT responders and non-responders are small, requiring hundreds of patients to be appropriately powered. We highlighted that symptom variance is an important confound to interpreting cross-sectional network effects and may drive prior findings of increased baseline connectivity in treatment non-responders. The baseline mean and variance of depression sum and symptom scores fared better at predicting response than both overall network connectivity and individual symptom strength centrality.

Supplementary material. The supplementary material for this article can be found at <https://doi.org/10.1017/S0033291723001368>

Acknowledgements. The authors wish to thank the employees of SilverCloud Health, Amwell for providing administrative and data processing support.

Authors' contributions. CTL and CMG conceptualised the research; CTL performed data processing and both CTL and CMG performed data analyses. CTL, SWK, and CMG interpreted the results while CTL and CMG wrote the paper. CTL, SWK, JP, DR, and CMG advised on, edited, and produced the final manuscript.

Financial support. This work was supported by the PhD studentship of CTL funded by the Irish Research Council's Enterprise Partnership Scheme, in conjunction with SilverCloud Health, Amwell (Grant number: EPSPG/2020/8).

Competing interests. This study forms part of the PhD studentship of CTL, which is co-funded by SilverCloud Health, Amwell (the provider of the iCBT intervention in this study) and the Irish Research Council. DR is a current employee of and hold shares in SilverCloud Health, Amwell. JP was a former employee of SilverCloud Health, Amwell. CMG has no direct competing interests but is the primary supervisor of CTL and SWK. SWK has no direct competing interests.

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.

Note

¹ The subsampling analysis pertaining to network connectivity changes following treatment included 79 Responder subsamples and 79 Non-Responder subsamples due to the removal of $n = 1$ subsample network that had a non-positive definite covariance matrix, which is unstable.

References

Beard, C., Millner, A. J., Forgeard, M. J. C., Fried, E. I., Hsu, K. J., Treadway, M. T., ... Björgvinsson, T. (2016). Network analysis of depression and anxiety

- symptom relationships in a psychiatric sample. *Psychological Medicine*, 46(16), 3359–3369. <https://doi.org/10.1017/S0033291716002300>.
- Berlim, M. T., Richard-Devantoy, S., Dos Santos, N. R., & Turecki, G. (2021). The network structure of core depressive symptom-domains in major depressive disorder following antidepressant treatment: A randomized clinical trial. *Psychological Medicine*, 51(14), 2399–2413. <https://doi.org/10.1017/S0033291720001002>.
- Blanco, I., Contreras, A., Chaves, C., Lopez-Gomez, I., Hervas, G., & Vazquez, C. (2020). Positive interventions in depression change the structure of well-being and psychological symptoms: A network analysis. *The Journal of Positive Psychology*, 15(5), 623–628. <https://doi.org/10.1080/17439760.2020.1789696>.
- Borsboom, D. (2017). A network theory of mental disorders. *World Psychiatry*, 16(1), 5–13. <https://doi.org/10.1002/wps.20375>.
- Borsboom, D., & Cramer, A. O. J. (2013). Network analysis: An integrative approach to the structure of psychopathology. *Annual Review of Clinical Psychology*, 9, 91–121. <https://doi.org/10.1146/annurev-clinpsy-050212-185608>.
- Bos, E. H., & De Jonge, P. (2014). “Critical slowing down in depression” is a great idea that still needs empirical proof. *Proceedings of the National Academy of Sciences*, 111(10), E878–E878. <https://doi.org/10.1073/pnas.1323672111>.
- Bos, F. M., Fried, E. I., Hollon, S. D., Bringmann, L. F., Dimidjian, S., DeRubeis, R. J., & Boeking, C. L. H. (2018). Cross-sectional networks of depressive symptoms before and after antidepressant medication treatment. *Social Psychiatry and Psychiatric Epidemiology*, 53(6), 617–627. <https://doi.org/10.1007/s00127-018-1506-1>.
- Bringmann, L. F. (2021). Person-specific networks in psychopathology: Past, present, and future. *Current Opinion in Psychology*, 41, 59–64. <https://doi.org/10.1016/j.copsyc.2021.03.004>.
- Burger, J., Isvoranu, A.-M., Lunansky, G., Haslbeck, J. M. B., Epskamp, S., Hoekstra, R. H., ... Blanken, T. (2022). Reporting standards for psychological network analyses in cross-sectional data. *Psychological Methods*. Advance online publication. <https://doi.org/10.1037/met0000471>.
- Chen, J., & Chen, Z. (2008). Extended Bayesian information criteria for model selection with large model spaces. *Biometrika*, 95(3), 759–771. <https://doi.org/10.1093/biomet/asn034>.
- Contreras, A., Nieto, I., Valiente, C., Espinosa, R., & Vazquez, C. (2019). The study of psychopathology from the network analysis perspective: A systematic review. *Psychotherapy and Psychosomatics*, 88(2), 71–83. <https://doi.org/10.1159/000497425>.
- Cramer, A. O. J., van Borkulo, C. D., Giltay, E. J., van der Maas, H. L. J., Kendler, K. S., Scheffer, M., & Borsboom, D. (2016). Major depression as a complex dynamic system. *PLOS ONE*, 11(12), e0167490. <https://doi.org/10.1371/journal.pone.0167490>.
- Cuijpers, P., Karyotaki, E., Ciharova, M., Miguel, C., Noma, H., & Furukawa, T. A. (2021). The effects of psychotherapies for depression on response, remission, reliable change, and deterioration: A meta-analysis. *Acta Psychiatrica Scandinavica*, 144(3), 288–299. <https://doi.org/10.1111/acps.13335>.
- Curtiss, J. E., Wallace, B., Fisher, L. B., Nyer, M., Jain, F., Cusin, C., & Pedrelli, P. (2021). Change processes in cognitive behavioral therapy and motivational interviewing for depression and heavy alcohol use: A network approach. *Journal of Affective Disorders Reports*, 3, 100040. <https://doi.org/10.1016/j.jadr.2020.100040>.
- Elliott, H., Jones, P. J., & Schmidt, U. (2020). Central symptoms predict post-treatment outcomes and clinical impairment in anorexia nervosa: A network analysis. *Clinical Psychological Science*, 8(1), 139–154. <https://doi.org/10.1177/2167702619865958>.
- Epskamp, S., Cramer, A. O. J., Waldorp, L. J., Schmittmann, V. D., & Borsboom, D. (2012). Qgraph: Network visualizations of relationships in psychometric data. *Journal of Statistical Software*, 48(1), 1–18. <https://doi.org/10.18637/jss.v048.i04>.
- Epskamp, S., & Fried, E. I. (2018). A tutorial on regularized partial correlation networks. *Psychological Methods*, 23(4), 617–634. <https://doi.org/10.1037/met0000167>.
- Epskamp, S., Waldorp, L. J., Möttus, R., & Borsboom, D. (2018). The Gaussian Graphical Model in cross-sectional and time-series data. *Multivariate Behavioral Research*, 53(4), 453–480. <https://doi.org/10.1080/00273171.2018.1454823>.
- Fisher, A. J., Medaglia, J. D., & Jeronimus, B. F. (2018). Lack of group-to-individual generalizability is a threat to human subjects research. *Proceedings of the National Academy of Sciences*, 115(27), E6106–E6115. <https://doi.org/10.1073/pnas.1711978115>.
- Forbes, M. K., Wright, A. G. C., Markon, K. E., & Krueger, R. F. (2017). Evidence that psychopathology symptom networks have limited replicability. *Journal of Abnormal Psychology*, 126(7), 969–988. <https://doi.org/10.1037/abn0000276>.
- Fried, E. I., van Borkulo, C. D., Epskamp, S., Schoevers, R. A., Tuerlinckx, F., & Borsboom, D. (2016). Measuring depression over time ... Or not? Lack of unidimensionality and longitudinal measurement invariance in four common rating scales of depression. *Psychological Assessment*, 28(11), 1354–1367. <https://doi.org/10.1037/pas0000275>.
- Friedman, E. S., Davis, L. L., Zisook, S., Wisniewski, S. R., Trivedi, M. H., Fava, M., ... Team, C. S. (2012). Baseline depression severity as a predictor of single and combination antidepressant treatment outcome: Results from the CO-MED trial. *European Neuropsychopharmacology*, 22(3), 183. <https://doi.org/10.1016/j.euroneuro.2011.07.010>.
- Hagan, K. E., Matheson, B. E., Datta, N., L’Insalata, A. M., Onipede, Z. A., Gorrell, S., ... Lock, J. D. (2021). Understanding outcomes in family-based treatment for adolescent anorexia nervosa: A network approach. *Psychological Medicine*, 53(2), 396–407. <https://doi.org/10.1017/S0033291721001604>.
- Hamaker, E. (2012). Why researchers should think ‘within-person’: A paradigmatic rationale. In M. R. Mehl & T. S. Conner (Eds.), *Handbook of research methods for studying daily life* (pp. 43–61). New York, NY: The Guilford Press.
- Harrison, P., Walton, S., Fennema, D., Duan, S., Jaekle, T., Goldsmith, K., ... Zahn, R. (2021). Development and validation of the Maudsley Modified Patient Health Questionnaire (MM-PHQ-9). *BJPsych Open*, 7(4), e123. <https://doi.org/10.1192/bjo.2021.953>.
- Ho, D., Imai, K., King, G., & Stuart, E. A. (2011). MatchIt: Nonparametric preprocessing for parametric causal inference. *Journal of Statistical Software*, 42(8), 1–28. <https://doi.org/10.18637/jss.v042.i08>.
- Jacobson, N. S., & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology*, 59(1), 12–19. <https://doi.org/10.1037/0022-006x.59.1.12>.
- Karyotaki, E., Efthimiou, O., Miguel, C., Bermpohl, F. M. G., Furukawa, T. A., Cuijpers, P., ... Forsell, Y. (2021). Internet-based cognitive behavioral therapy for depression: A systematic review and individual patient data network meta-analysis. *JAMA Psychiatry*, 78(4), 361–371. <https://doi.org/10.1001/jamapsychiatry.2020.4364>.
- Kelley, S. W., Fisher, A. J., Lee, C. T., Gallagher, E., Hanlon, A. K., Robertson, I., ... Gillan, C. M. (2022). Elevated emotion network connectivity leaves people vulnerable to fluctuations in depression [Preprint]. *PsyArXiv*. Accessed on <https://psyarxiv.com/agt5m/>.
- Kroenke, K., Spitzer, R. L., & Williams, J. B. W. (2001). The PHQ-9: Validity of a brief depression severity measure. *Journal of General Internal Medicine*, 16(9), 606–613. <https://doi.org/10.1046/j.1525-1497.2001.016009606.x>.
- Lawler, K., Earley, C., Timulak, L., Enrique, A., & Richards, D. (2021). Dropout from an internet-delivered cognitive behavioral therapy intervention for adults with depression and anxiety: Qualitative study. *JMIR Formative Research*, 5(11), e26221. <https://doi.org/10.2196/26221>.
- Linn, R. L. (1968). Range restriction problems in the use of self-selected groups for test validation. *Psychological Bulletin*, 69(1), 69–73. <https://doi.org/10.1037/h0025263>.
- Maj, M., Stein, D. J., Parker, G., Zimmerman, M., Fava, G. A., De Hert, M., ... Wittchen, H. U. (2020). The clinical characterization of the adult patient with depression aimed at personalization of management. *World Psychiatry*, 19(3), 269–293. <https://doi.org/10.1002/wps.20771>.
- McElroy, E., Napoleone, E., Wolpert, M., & Patalay, P. (2019). Structure and connectivity of depressive symptom networks corresponding to early treatment response. *EClinicalMedicine*, 8, 29–36. <https://doi.org/10.1016/j.eclinm.2019.02.009>.

- McMahon, F. J. (2014). Prediction of treatment outcomes in psychiatry – Where do we stand? *Dialogues in Clinical Neuroscience*, 16(4), 455. <https://doi.org/10.31887/DCNS.2014.16.4/FMCMAHON>.
- O'Driscoll, C., Buckman, J. E. J., Fried, E. I., Saunders, R., Cohen, Z. D., Ambler, G., ... Pilling, S. (2021). The importance of transdiagnostic symptom level assessment to understanding prognosis for depressed adults: Analysis of data from six randomised control trials. *BMC Medicine*, 19(1), 109. <https://doi.org/10.1186/s12916-021-01971-0>.
- Palacios, J. E., Enrique, A., Mooney, O., Farrell, S., Earley, C., Duffy, D., ... Richards, D. (2022). Durability of treatment effects following internet-delivered cognitive behavioural therapy for depression and anxiety delivered within a routine care setting. *Clinical Psychology & Psychotherapy*, 29(5), 1768–1777. <https://doi.org/10.1002/cpp.2743>.
- Papini, S., Rubin, M., Telch, M. J., Smits, J. A. J., & Hien, D. A. (2020). Pretreatment posttraumatic stress disorder symptom network metrics predict the strength of the association between node change and network change during treatment. *Journal of Traumatic Stress*, 33(1), 64–71. <https://doi.org/10.1002/jts.22379>.
- Pe, M. L., Kircanski, K., Thompson, R. J., Bringmann, L. F., Tuerlinckx, F., Mestdagh, M., ... Gotlib, I. H. (2015). Emotion-network density in major depressive disorder. *Clinical Psychological Science*, 3(2), 292–300. <https://doi.org/10.1177/2167702614540645>.
- Richards, D., Enrique, A., Eilert, N., Franklin, M., Palacios, J., Duffy, D., ... Timulak, L. (2020). A pragmatic randomized waitlist-controlled effectiveness and cost-effectiveness trial of digital interventions for depression and anxiety. *Npj Digital Medicine*, 3(85), 1–10. <https://doi.org/10.1038/s41746-020-0293-8>.
- Robinaugh, D. J., Hoekstra, R. H. A., Toner, E. R., & Borsboom, D. (2020). The network approach to psychopathology: A review of the literature 2008–2018 and an agenda for future research. *Psychological Medicine*, 50(3), 353–366. <https://doi.org/10.1017/S0033291719003404>.
- Robinaugh, D. J., Millner, A. J., & McNally, R. J. (2016). Identifying highly influential nodes in the complicated grief network. *Journal of Abnormal Psychology*, 125(6), 747–757. <https://doi.org/10.1037/abn0000181>.
- Rodebaugh, T. L., Tonge, N. A., Piccirillo, M. L., Fried, E., Horenstein, A., Morrison, A. S., ... Heimberg, R. G. (2018). Does centrality in a cross-sectional network suggest intervention targets for social anxiety disorder? *Journal of Consulting and Clinical Psychology*, 86(10), 831–844.
- Rost, N., Binder, E. B., & Brückl, T. M. (2023). Predicting treatment outcome in depression: An introduction into current concepts and challenges. *European Archives of Psychiatry and Clinical Neuroscience*, 273(1), 113–127. <https://doi.org/10.1007/s00406-022-01418-4>.
- Rush, A. J., Trivedi, M. H., Wisniewski, S. R., Nierenberg, A. A., Stewart, J. W., Warden, ... Fava, M. (2006). Acute and longer-term outcomes in depressed outpatients requiring one or several treatment steps: A STAR*D report. *The American Journal of Psychiatry*, 163(11), 1905–1917. <https://doi.org/10.1176/AJP.2006.163.11.1905>.
- Schumacher, L., Burger, J., Echterhoff, J., & Kriston, L. (2022). Methodological and Statistical Practices of using Symptom Networks to Evaluate Mental Health Interventions: A Systematic Review [Preprint]. *PsyArXiv*. Accessed on <https://psyarxiv.com/ea7fu>.
- Schweren, L., van Borkulo, C. D., Fried, E., & Goodyer, I. M. (2018). Assessment of symptom network density as a prognostic marker of treatment response in adolescent depression. *JAMA Psychiatry*, 75(1), 98–100. <https://doi.org/10.1001/jamapsychiatry.2017.3561>.
- Spiller, T. R., Levi, O., Neria, Y., Suarez-Jimenez, B., Bar-Haim, Y., & Lazarov, A. (2020). On the validity of the centrality hypothesis in cross-sectional between-subject networks of psychopathology. *BMC Medicine*, 18(1), 1–14. <https://doi.org/10.1186/s12916-020-01740-5>.
- Spitzer, R. L., Kroenke, K., Williams, J. B. W., & Löwe, B. (2006). A brief measure for assessing generalized anxiety disorder. *Archives of Internal Medicine*, 166(10), 1092. <https://doi.org/10.1001/archinte.166.10.1092>.
- Terluin, B., de Boer, M. R., & de Vet, H. C. W. (2016). Differences in connection strength between mental symptoms might be explained by differences in variance: Reanalysis of network data did not confirm staging. *PLOS ONE*, 11(11), e0155205. <https://doi.org/10.1371/journal.pone.0155205>.
- Trivedi, M. H., Rush, A. J., Wisniewski, S. R., Nierenberg, A. A., Warden, D., Ritz, L., ... Fava, M. (2006). Evaluation of outcomes with citalopram for depression using measurement-based care in STAR*D: Implications for clinical practice. *American Journal of Psychiatry*, 163(1), 28–40. <https://doi.org/10.1176/appi.ajp.163.1.28>.
- van Borkulo, C. D., Boschloo, L., Borsboom, D., Penninx, B. W. J. H., Lourens, J. W., & Schoevers, R. A. (2015). Association of symptom network structure with the course of longitudinal depression. *JAMA Psychiatry*, 72(12), 1219–1226. <https://doi.org/10.1001/jamapsychiatry.2015.2079>.
- van Borkulo, C. D., van Bork, R., Boschloo, L., Kossakowski, J. J., Tio, P., Schoevers, R. A., ... Waldorp, L. J. (2022). Comparing network structures on three aspects: A permutation test. *Psychological Methods*. Advance online publication. <https://doi.org/10.1037/met0000476>.
- von Klipstein, L., Borsboom, D., & Arntz, A. (2021). The exploratory value of cross-sectional partial correlation networks: Predicting relationships between change trajectories in borderline personality disorder. *PLOS ONE*, 16(7), e0254496. <https://doi.org/10.1371/journal.pone.0254496>.
- Wing, J. K., Babor, T., Brugha, T., Burke, J., Cooper, J. E., Giel, R., ... Sartorius, N. (1990). SCAN: Schedules for Clinical Assessment in Neuropsychiatry. *Archives of General Psychiatry*, 47(6), 589–593. <https://doi.org/10.1001/archpsyc.1990.01810180089012>.
- World Health Organization. (2023). *Depression*. Retrieved from <https://www.who.int/news-room/fact-sheets/detail/depression>.