Early Change in Specific Depression Symptoms and Later Outcome in Internet-Delivered Psychotherapy for Depression: A Cohort Study and Cross-Lagged Network Analysis

Fred Johansson[1,2], Oskar Flygare[1], Julia Bäckman[1], Robin Fondberg[1], Erland Axelsson [3,4], Erik Forsell[1], Matti Cervin [5], Viktor Kaldo[1, 6], Christian Rück[1] & John Wallert[1]

1. Centre for Psychiatry Research, Department of Clinical Neuroscience, Karolinska Institutet, & Stockholm Health Care Services, Region Stockholm, Stockholm, Sweden
2. Department of Health Promotion Science, Sophiahemmet University, Sweden
3. Division of Family Medicine and Primary Care, Department of Neurobiology, Care Sciences and Society, Karolinska Institutet, Stockholm, Sweden
4. Liljeholmen University Primary Health Care Centre, Academic Primary Health Care Centre, Region Stockholm, Stockholm, Sweden
5. Department of Clinical Sciences, Lund University, Lund, Sweden
6. Department of Psychology, Faculty of Health and Life Sciences, Linnaeus University, Växjö, Sweden

Corresponding author: Fred Johansson, fredrik.johansson@ki.se

Word count: 4490

Abstract

Background. Symptom reduction occurring early in depression treatment is associated with favourable post-treatment outcome, but it is not known how early reduction in *specific* depression symptoms affect treatment outcome. We aimed to determine the impact of symptom-specific change from pre-treatment to week four during internet-delivered CBT (ICBT) on overall and symptom-specific depression severity at post-treatment. We hypothesized that change in mood and emotional involvement would be most strongly associated with later overall depression severity.

Methods. 1300 participants with Major Depressive Disorder were followed over 12 weeks of ICBT using the self-report Montgomery-Åsberg Depression Rating Scale gauging nine symptoms. Linear models, informed by causal inference and cross-lagged network analysis methods, were used to estimate associations between early symptom-specific change and post-treatment depression severity, controlling for register-based and self-reported pre-treatment confounders.

Results. Early reduction in all symptoms was associated with lower overall and symptom-specific depression severity post-ICBT. Seven symptoms showed similar associations between early change and overall depression severity post-treatment: mood (standardized beta [β]=0.44), feelings of unease (β=0.39), ability to concentrate (β=0.46), initiative (β=0.43), emotional involvement (β=0.42), pessimism (β=0.44), and zest for life (β=0.42). Change in sleep (β=0.27) and appetite (β=0.27) had weaker associations with overall depression severity at post-treatment and were the only symptoms showing the hypothesized difference compared with mood and emotional involvement.

Conclusions. The impact of early symptom-specific reduction on post-treatment depression severity in ICBT for MDD may be similar across most symptoms, but less for the sleep and appetite symptoms, although causal interpretations rests on several assumptions.

Introduction

Symptom reduction during the first few weeks of Major Depressive Disorder (MDD) treatment is associated with higher rates of remission and lower symptom burden post-treatment across psychological and medical therapies (Beard & Delgadillo, 2019; Solmi et al., 2023; Szegedi et al., 2009). This includes internet-delivered cognitive behavioural therapy (ICBT), where reduced symptoms during the first 4-6 weeks is predictive of a better post-treatment outcome (Forsell et al., 2020; Schibbye et al., 2014).

Depression, however, is a heterogeneous disorder with numerous possible symptom combinations (Fried & Nesse, 2015; Maj et al., 2020). Overall improvement during treatment is therefore not necessarily equally distributed across specific symptoms (Garke et al., 2021; Kopta et al., 1994). It is not known whether early improvement in some *specific* depression symptoms is more important for treatment outcome than improvement in others. Identifying specific influential symptoms could indicate specific symptoms to target in treatment. Such guidance is highly needed, given that knowledge on key components of psychological treatments for depression is very limited (Cuijpers et al., 2019).

Shifting the focus from overall depression scores to individual symptoms is in line with the network theory of mental disorders (Borsboom, 2017). This theory propose that mental disorders are the result bi-directional causal influences between symptoms. Change in a specific symptom is assumed to influence other symptoms in the network, leading to effects that spread through the symptom network and affect overall symptom severity. Accordingly, a central task of network analysis is to identify the most influential symptoms in the network, which could inform treatment design as effectively targeting influential symptoms could have widespread effects across the symptom network (Borsboom & Cramer, 2013). The influence of different symptoms in a network is usually quantified through estimating centrality, indicating how strongly a specific symptom is correlated with other symptoms in the network. Mood and emotional involvement are considered hallmark symptoms of depression in DSM-5 (American Psychiatric Association, 2013) and ICD-11 (World Health Organization, n.d.), and have been identified as the most central symptoms of depression in cross-sectional network analyses, along with energy loss and concentration problems (Wichers et al., 2021). Mood and anhedonia were also the two depression symptoms most strongly associated with MDD onset over a six-year follow-up in a population of adults without a prior history of depression (Boschloo et al., 2016).

A limitation of much prior network theory research is that, although the theory concerns causal interactions between symptoms, it has predominantly been evaluated using cross-sectional designs that provide weak evidence of causality (McNally, 2021). Longitudinal data provide the opportunity for stronger causal inference, making it possible to disentangle temporal relationships between symptoms by adjusting for prior symptom levels and potential confounders (VanderWeele et al., 2016). Here we combined modern epidemiological methods for causal inference (VanderWeele, 2022; VanderWeele et al., 2016) with the recently proposed cross-lagged network model (Wysocki et al., 2022).

We aimed to determine the impact of early change in specific depression symptoms during the first four weeks of ICBT on overall depression severity at post-treatment. Further, we aimed to explore the impact of early change in each depression symptom on symptom-specific levels at post-treatment (i.e. a cross-lagged network analysis). We hypothesized that early change in the two hallmark symptoms of depression - mood and emotional involvement (anhedonia) - would be more strongly associated with overall depression severity at post-treatment than change in other symptoms.

Methods

Design and participants

This study leveraged the new MULTI-PSYCH cohort (Boberg et al., 2023), containing multi-modal data (clinical, registry and genetic) for adults treated with ICBT at the Internet Psychiatry Unit in Stockholm. Participants were recruited from 2009 through 2019. We included the 1,300 participants who were diagnosed with mild-to-moderate MDD and enrolled in a 12-week ICBT intervention for MDD (Hedman et al., 2014). Self-rated depression symptoms were collected weekly via an online treatment platform and the present analyses used data from pre-treatment, week four of treatment, and post-treatment (week 12). Additional data were taken from several linked Swedish administrative registers including the National Patient Register, the Prescribed Drug Register and the Longitudinal integrated database for health insurance and labour market studies. Detailed information on the MULTI-PSYCH cohort, and the ICBT treatment for MDD, is available elsewhere (Boberg et al., 2023). This study was approved by the Regional Ethics Board in Stockholm and the National Ethical Review Authority (approval numbers 2009/1089- 31/2, 2022-00602-02) and all participants provided informed consent before entering the study.

Measures

The self-reported Montgomery-Åsberg Depression Rating Scale (MADRS-S) (Svanborg & Åsberg, 1994) was used to measure both specific depression symptoms and overall depression severity. MADRS-S comprises nine items rated on a 0-6 response scale, with higher numbers indicating more severe symptoms. MADRS-S includes items capturing the following symptoms within a three-day recall period: 1) mood, 2) feelings of unease, 3) (reduced) sleep, 4) (reduced) appetite, 5) ability to concentrate, 6) initiative, 7) emotional involvement (anhedonia), 8) pessimism and 9) zest for life. Item responses are summed for a total score of 0-54, which was used as our overall depression severity measure, while items specific ratings were used as measures of specific symptoms. The MADRS-S has shown adequate psychometric properties for measuring depression when conceptualised as a latent construct (van Ballegooijen et al., 2016). Cronbach's alpha in the current sample was 0.78 at pre-treatment, 0.88 at week four and 0.92 at post-treatment.

The Patient Health Questionnaire (PHQ-9) (Kroenke et al., 2001) was used as an alternative self-rated measure of overall depression severity in sensitivity analyses. PHQ-9 comprises nine items rated from 0 (Not at all) to 3 (Nearly every day), intended to correspond to the DSM criteria for MDD. The items are summed for a total score of 0-27. Cronbach's alpha of PHQ-9 at post-treatment was 0.87 in the current sample.

Covariates. We adjusted for the following pre-treatment covariates to reduce confounding: depression symptoms (each specific items of MADRS-S), sex (male/female), age (years), highest attained education level (primary/secondary/higher), disposable income in the year prior treatment (categorized into quintiles), civil status (married/unmarried/separated), self-reported prior diagnosis of MDD (no/mild/moderate or severe), prescription of anti-depressant medication in the year before treatment (yes/no), self-reported prior psychiatric in-patient care (yes/no), self-reported prior suicide attempt (yes/no), self-reported alcohol risk score, version of ICBT protocol, self-reported insomnia score, psychiatric comorbidity in the year prior to treatment (yes/no), days of full or partial paid sick leave in the year prior to treatment (≤ 14/>14), primary employment status in the year prior to treatment

(employed/unemployed). Covariates were collected from self-reports and from linked Swedish administrative registers. Detailed definitions and sources of the covariates are provided in Supplementary table 1.

Statistical analyses

Descriptive statistics for the sample are presented as mean (SD) or count (%), as appropriate (Table 1). The individual symptom scores, as well as the total MADRS-S score were standardized (mean=0, SD=1) before analysis.

Change in specific depression symptoms and overall depression severity at post-treatment

The association between change in each of the depression symptoms from pre-treatment to week four and overall depression severity at post-treatment was estimated by regressing the MADRS-S sum score at post-treatment on each of the individual depression symptoms at week four. We built separate linear regression models for each symptom at week four with adjustment for all specific pre-treatment symptoms along with the other pre-treatment covariates (nine models in total). Formal model definitions are provided in the Supplementary Methods. Multi-collinearity among the symptoms was checked using the variance inflation factor that was <2 in all complete case models. Estimates are presented as standardized betas along with 95% confidence intervals in Table 2.

By controlling for pre-treatment symptom levels, the coefficient for each specific symptom at week four can be interpreted as the association between a *change* in that specific symptom from pre-treatment to week four and MADRS-S score at post-treatment, adjusted for all other symptoms at pre-treatment (for details, see Supplementary Methods 1 and the eMethods2 in Nakamura et al. (2022)). Controlling for each specific symptom at pre-treatment provides stronger adjustment for potential confounding and reverse causality (VanderWeele et al., 2016), and has been proposed as a method to estimate the effects of separate items of a scale (VanderWeele, 2022). Change in the other symptoms were not adjusted for (i.e. only the symptom of interest at week four was included in the model), as these were assumed to be on the causal path from change in a specific symptom and post-treatment depression severity. The modelling approach was guided by the assumed causal structure depicted in the directed acyclic graph in Figure 1 (see Supplemental Methods 2 for an elaboration).

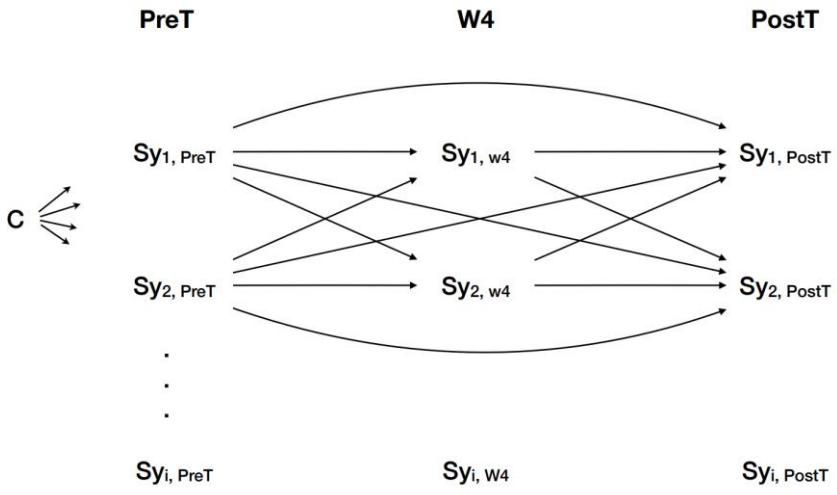

Figure 1. Directed acyclic graph showing the assumed causal structure underlying the models. Sy_i denotes the individual symptoms at time pre-treatment (PreT), week 4 of treatment (W4) and post-treatment (PostT). Only

symptom 1 and 2 are shown to avoid clutter, but the same structure is assumed for all the nine symptoms. C denotes the set of pre-treatment covariates.

We hypothesized that early change in the each of the two hallmark symptoms of depression, mood and emotional engagement, would be more strongly associated with overall depression severity at post-treatment than early change in the other symptoms. This was tested by comparing the coefficients for the mood and emotional involvement symptoms against the coefficients for the seven other symptoms using two-sided z-tests at $\alpha = 0.05$ (Paternoster et al., 1998) (Table 2).

Cross-lagged network analysis

Similar models were used to estimate associations between symptom-specific change and post-treatment levels *of specific symptoms*. We fit separate models for each symptom (week four) – symptom (post-treatment) combination, yielding 81 models that where were controlled for all pre-treatment symptoms as well as the set of other pre-treatment covariates. A formal definition of these models is provided in the Supplementary Methods 1. The modelling approach was inspired by the Cross-Lagged Network Model proposed by Wysocki et al. (2022) but was modified to control for prior symptom levels, for reasons described above. Estimated associations are presented as standardized betas with 95% confidence intervals in Table 3 and visualised as a network in Figure 2.

Expected influence

The one-step expected influence (EI) of each symptom was calculated as the sum of all edges going out of a symptom (including the auto-regressive association that a symptom has with itself at a later time-point) and are presented in Table 3. This is a modification of the cross-sectional EI that sums all edges *connected* to a node, but excludes the auto-correlation which is always 1 in cross-sectional analyses (Robinaugh et al., 2016).

Missing data

Missing data was handled by using multiple imputation by chained equations to create 20 imputed datasets. Analyses was performed on each of the 20 imputed dataset and estimates were pooled using Rubin's rules. More information on the multiple imputation procedure is provided in Supplementary Methods 3, with imputation diagnostics presented in Supplementary Figures 2 and 3. We also conducted complete case analyses restricted to participants having MADRS-S data at pre-treatment, week four and post-treatment, adjusted only for pre-treatment symptom levels (Supplementary Table 2).

Sensitivity analyses

We conducted several sensitivity analyses. First, we performed a sensitivity analysis using the PHQ-9 sum score at post-treatment as the outcome (instead of MADRS-S) to assess the sensitivity of our estimates to the depression measure used (Supplementary Table 3).

Second, we performed sensitivity analyses with MADRS-S sum scores excluding the predictor symptom as outcome (e.g. by regressing the sum of symptom 2-9 at post-treatment on symptom 1 at week four) to assess the extent to which auto-regressive effects could explain associations (Supplementary Table 3).

Third, potential effect modification by initial depression severity was assessed by stratifying the sample by the median of MADRS-S severity at pre-treatment and performing analyses in each of the strata separately (Supplementary Table 4).

Fourth, to partially account for instability in the individual item measurements, we performed sensitivity analyses using individual regression slopes of the weekly measures of each symptom from pre-treatment to week four as predictor variables. Slopes were obtained by regressing weekly values of each symptom from pre-treatment to week four on time for each participant individually (Supplementary Table 5).

Finally, to investigate whether differences in connection strength could be explained by differences in mean or variance of the symptoms (Lee et al., 2023; Terluin et al., 2016) we plotted the association of each symptom at week four and depression severity at post-treatment (standardized betas) against their mean and SD at week four (Supplementary Figure 3).

A pre-registration of this analytic plan and the R code used is available at: https://osf.io/7ga8k/. The only changes in relation to the pre-registration were slight redefinitions of the covariates psychiatric comorbidity, prior MDD episodes and disposable income (Supplementary Table 1). All analyses were performed in R version 4.3.2.

Results

We included 1300 individuals diagnosed with MDD who were followed over a 12-week ICBT treatment for depression. The mean age was 37.6 years (SD 11.9), 66% were female and 42% has been prescribed antidepressant medication in the year prior to treatment (Table 1). The mean MADRS-S score was 22.7 (SD 6.3) at pre-treatment, 17.1 (7.5) at week four, and 13.0 (SD 8.1) at post-treatment. The median duration between pre-treatment and week four was 28 days (interquartile range = 2). Symptom-specific mean (SD) levels at week four were: mood 1.78 (1.15), feelings of unease 2.41 (1.20), sleep 1.95 (1.33), appetite 0.89 (1.07), ability to concentrate 2.05 (1.17), initiative 2.11 (1.19), emotional involvement 1.85 (1.00), pessimism 2.57 (1.35), zest for life 1.52 (0.98). Data on the MADRS-S was missing for 14 (1%) participants at pre-treatment, 227 (17%) at week four and 248 (19%) at post-treatment. Missingness for the covariates is presented in Table 1.

Table 1. Pre-treatment characteristics of the full sample		
	n = 1300	NA, n (%)
Sociodemographic		
Age	37.63 (11.87)	0 (0%)
Male	443 (34%)	0 (0%)
Civil status		5 (0%)
Married	336 (26%)	
Unmarried	806 (62%)	
Separated	153 (12%)	
Education level		5 (0%)
Primary	64 (5%)	
Secondary	365 (28%)	
Higher	866 (67%)	
Disposable income [a]		5 (0%)
1st quintile	68.48 (41.13)	
2nd quintile	169.31 (20.79)	
3rd quintile	232.31 (16.68)	

4th quintile	295.10 (23.57)	
5th quintile	515.74 (298.90)	
Unemployed	167 (18%)	355 (27%)
Sick leave ≥14 days	156 (12%)	5 (0%)
Clinical history		
Psychiatric comorbidity	134 (10%)	0 (0%)
Prior MDD episodes		47 (4%)
No	760 (61%)	
Mild	213 (17%)	
Moderate/severe	280 (22%)	
Antidepressant medication in the year before treatment	551 (42%)	0 (%)
Prior psychiatric in-patient care	90 (7%)	55 (4%)
Prior suicide attempt	76 (6%)	95 (7%)
Self-rated scales		
Alcohol risk score	5.42 (4.76)	50 (4%)
Insomnia score	12.73 (5.86)	18 (1%)
MADRS-S total score	22.67 (6.27)	14 (1%)
MADRS-S individual items		14 (1%)
Mood	2.51 (1.11)	14 (1%)
Feelings of unease	3.08 (1.10)	14 (1%)
Sleep	2.30 (1.40)	14 (1%)
Appetite	1.24 (1.28)	14 (1%)
Ability to concentrate	2.60 (1.18)	14 (1%)
Initiative	2.97 (1.19)	14 (1%)
Emotional involvement (Anhedonia)	2.52 (1.04)	14 (1%)
Pessimism	3.44 (1.18)	14 (1%)
Zest for life	2.01 (1.02)	14 (1%)

Data are integer count (%) or decimal mean (SD).
Definitions of the variables are provided in Supplementary Table 1.
[a] Disposable income is used as a categorical variable in analyses, but here presented as family-adjusted disposable income in 1000 SEK.

Change in individual depression symptoms and overall depression severity at post-treatment

Change in all nine depression symptoms from pre-treatment to week four were positively associated with overall depression severity at post-treatment. Most of these associations were of similar strength with standardized betas between 0.39 and 0.46 (Table 2). The two exceptions were the change in sleep problems and decreased appetite, which showed weaker associations with overall depression severity at post-treatment: both with a standardized betas of 0.27 (Table 2). We hypothesised that change in the two hallmark symptoms of depression, mood and emotional engagement, would be more strongly associated with overall depression severity at post-treatment, compared to change in other symptoms. This was the case in relation to sleep and appetite that showed significantly weaker associations to post-treatment depression severity compared with both mood and emotional engagement. It was however

not the case in relation to the other symptoms: feelings of unease, ability to concentrate, initiative pessimism or zest for life, whose associations to post-treatment depression severity were not significantly different from either mood or emotional involvement (Table 2). Complete case analyses with 942 participants gave similar results, with standardized betas within 0.02 from the main analysis (Supplementary Table 2).

Table 2. Associations between change in individual symptoms between pre-treatment and week four and overall depression levels at post-treatment

Symptom	**Standardized beta (95% CI)** [a]
Mood	0.44 (0.38-0.51)
Feelings of unease	0.39 (0.32-0.46)
Sleep	0.27 (0.19-0.36) [b,c]
Appetite	0.27 (0.20-0.34) [b,c]
Ability to concentrate	0.46 (0.39-0.53)
Initiative	0.43 (0.36-0.49)
Emotional involvement (Anhedonia)	0.42 (0.36-0.49)
Pessimism	0.44 (0.37-0.52)
Zest for life	0.42 (0.35-0.50)

[a] Adjusted for pre-treatment levels of: all nine individual MADRS-S symptoms, age, sex, education level, disposable income, unemployment, sick-leave, psychiatric comorbidity, prior MDD episodes, anti-depressant medication, psychiatric in-patient care, suicide attempts, version of ICBT protocol, alcohol risk score and insomnia severity.
[b] Significantly weaker association compared to mood at $p<0.05$
[c] Significantly weaker association compared to emotional involvement at $p<0.05$

Cross-lagged network analysis

The cross-lagged network analysis showed that change in all symptoms from pre-treatment to week four of ICBT were positively associated with symptom levels for all individual symptoms at post-treatment (Table 3, Figure 2). Standardized betas of auto-regressive associations ranged 0.40-0.52. For the cross-symptom associations, most standardized betas were around 0.30-0.40, except for those associations including the sleep or appetite symptoms. Early change in sleep and appetite was associated with other individual symptoms at post-treatment with standardized betas ranging 0.16-0.24. Similarly, change in the other seven symptoms were associated with sleep and appetite at post-treatment with standardized betas of 0.15-0.30. (Table 3, Figure 2). The expected influence was lowest for sleep (1.84) and appetite (1.89), while it showed largely similar levels for the other seven symptoms (2.67-3.18) (Table 3).

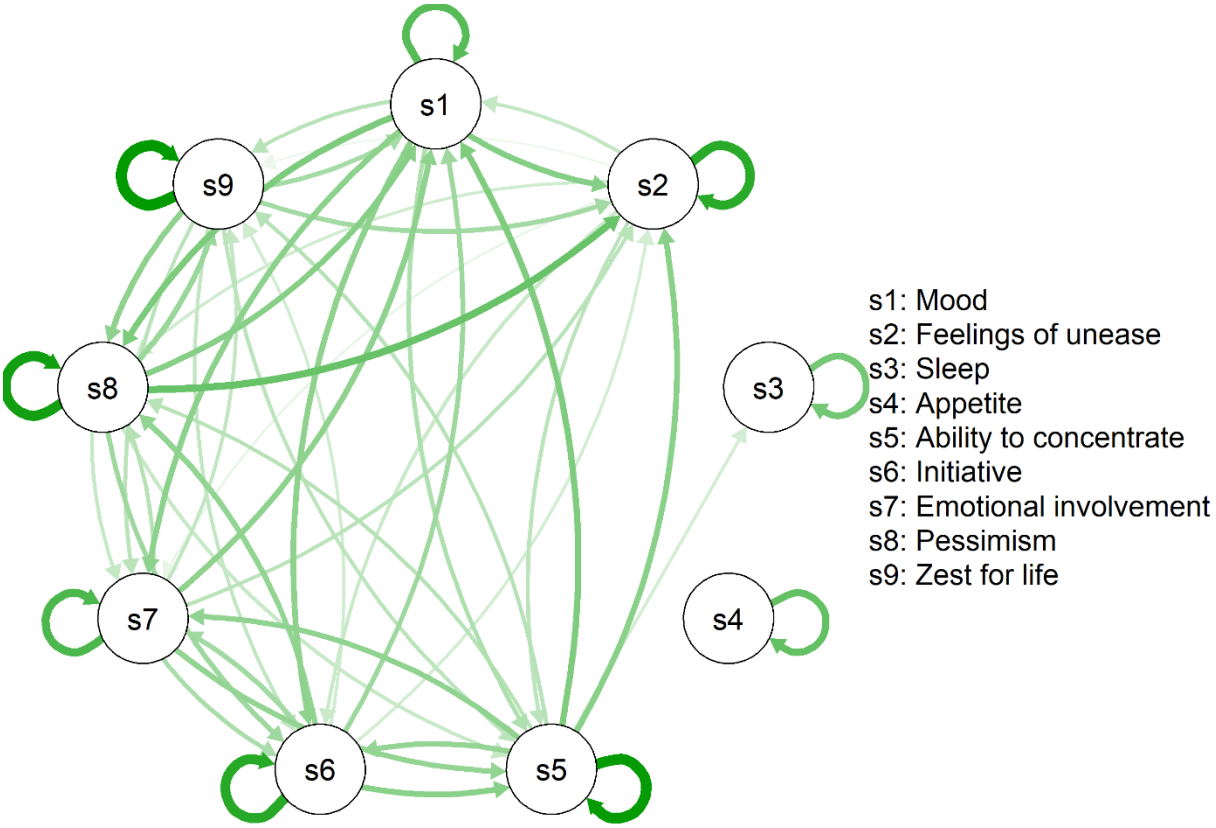

Figure 2. Cross-lagged network analysis with edges representing cross-lagged associations between symptom-specific change from pre-treatment to week four and symptom-specific levels at post-treatment. The width and saturation of edges are proportional to associational strength, edges with absolute values below 0.25 are not shown.

Table 3. Associations between change in individual symptoms from pre-treatment to week four and symptoms at post treatment, and the expected influence of each symptom

	Standardized beta [a] (95% CI)									EI
	S1-PostT	**S2-PostT**	**S3-PostT**	**S4-PostT**	**S5-PostT**	**S6-PostT**	**S7-PostT**	**S8-PostT**	**S9-PostT**	
S1-W4	0.43 (0.36-0.50)	0.38 (0.31-0.45)	0.22 (0.15-0.29)	0.22 (0.15-0.29)	0.34 (0.27-0.40)	0.37 (0.30-0.44)	0.37 (0.31-0.44)	0.39 (0.32-0.46)	0.33 (0.26-0.39)	3,05
S2-W4	0.31 (0.24-0.39)	0.48 (0.41-0.55)	0.20 (0.13-0.27)	0.20 (0.12-0.28)	0.32 (0.25-0.39)	0.31 (0.24-0.38)	0.27 (0.20-0.35)	0.30 (0.23-0.37)	0.27 (0.20-0.33)	2,67
S3-W4	0.17 (0.07-0.26)	0.22 (0.14-0.30)	0.40 (0.33-0.47)	0.12 (0.02-0.21)	0.24 (0.16-0.33)	0.17 (0.08-0.26)	0.16 (0.07-0.26)	0.21 (0.12-0.29)	0.16 (0.08-0.24)	1,84
S4-W4	0.19 (0.11-0.27)	0.22 (0.14-0.30)	0.20 (0.13-0.28)	0.42 (0.34-0.50)	0.18 (0.11-0.26)	0.18 (0.10-0.25)	0.16 (0.08-0.24)	0.16 (0.09-0.24)	0.17 (0.10-0.24)	1,89
S5-W4	0.39 (0.31-0.47)	0.38 (0.30-0.45)	0.30 (0.22-0.37)	0.21 (0.14-0.29)	0.52 (0.45-0.59)	0.36 (0.29-0.44)	0.37 (0.29-0.44)	0.33 (0.26-0.40)	0.33 (0.26-0.40)	3,18
S6-W4	0.35 (0.27-0.44)	0.30 (0.22-0.37)	0.19 (0.12-0.26)	0.21 (0.13-0.28)	0.38 (0.31-0.45)	0.48 (0.42-0.55)	0.35 (0.28-0.42)	0.37 (0.29-0.44)	0.30 (0.23-0.37)	2,93
S7-W4	0.37 (0.30-0.44)	0.33 (0.25-0.40)	0.23 (0.16-0.30)	0.21 (0.13-0.28)	0.37 (0.30-0.43)	0.34 (0.27-0.41)	0.44 (0.38-0.51)	0.33 (0.26-0.4)	0.32 (0.25-0.38)	2,93
S8-W4	0.38 (0.29-0.46)	0.41 (0.34-0.49)	0.18 (0.09-0.26)	0.22 (0.14-0.29)	0.31 (0.24-0.38)	0.35 (0.28-0.43)	0.31 (0.23-0.39)	0.50 (0.44-0.57)	0.36 (0.29-0.43)	3,02
S9-W4	0.35 (0.26-0.44)	0.35 (0.26-0.43)	0.15 (0.07-0.23)	0.25 (0.17-0.33)	0.32 (0.24-0.40)	0.32 (0.23-0.40)	0.32 (0.24-0.41)	0.37 (0.29-0.45)	0.52 (0.45-0.60)	2,96

Autoregressive associations are marked by dark grey shading while cross-symptom associations <0.25 are marked by light grey shading.
PostT: post-treatment; s1: mood; s2: feelings of unease; s3; sleep; s4: appetite; s5: ability to concentrate; s6: initiative; s7: emotional involvement; s8: pessimism; s9: zest for life; W4: week 4.

[a] Adjusted for pre-treatment levels of: the individual MADRS-S symptoms, age, sex, education level, disposable income, unemployment, sick-leave, psychiatric comorbidity, prior MDD episodes, anti-depressant medication, psychiatric in-patient care, suicide attempts, version of ICBT protocol, alcohol risk score and insomnia severity.

Sensitivity analyses

All sensitivity analyses were performed using overall depression severity at post-treatment as the outcome. First, we used the PHQ-9 total score as the outcome instead of MADRS-S, which gave similar but somewhat weaker, associations (standardized beta difference -0.02 to -0.08) (Supplementary Table 3).

Second, to exclude symptom-level auto-regressive effects, we also conducted analyses with each respective exposure symptoms excluded from the outcome MADRS-S sum score at post-treatment, again yielding similar results (standardized beta difference -0.01 to -0.04) (Supplementary Table 3).

Third, we conducted separate analyses for participants with high vs. low depression at pre-treatment to assess potential effect modification by depression severity. The sample was stratified using a median-split approach, where participants scoring above the median on pre-treatment MADRS-S (>23) were designated to the “high depression” group (n=588) and participants scoring at or below the median (≤23) were designated to the “low depression” group (n = 712). The pattern of associations was similar in both groups, but associations were stronger in the high depression group, with the appetite symptom showing

the most pronounced attenuation in the low depression group, and sleep showing the least attenuation (Supplementary Table 4).

Fourth, we regressed MADRS-S post-treatment score on each participants regression slope for each symptom from pre-treatment to week four, while adjusting for pre-treatment levels of the individual MARDS-S symptoms, but not the other covariates. This analysis was performed on a subset of the participants with complete data on the weekly MADRS-S ratings at pre-treatment, week 1-4 and post-treatment (n=777). Even though both the sample and adjustment set differed, results were very similar to the main analysis (absolute difference in standardized betas 0.00-0.06) (Supplementary Table 5).

Finally, plotting standardized betas against mean and SD at week four showed no clear pattern between strength of associations and the mean or SD of each symptom at week four (Supplementary Figure 4).

Discussion

This study aimed to determine the impact of symptom-specific change during the first four weeks of internet-delivered psychotherapy on post-treatment depression severity in a large sample of adults treated for major depressive disorder. Early reduction in each of the nine depression symptoms was associated with lower overall depression severity at post-treatment. Contrary to our hypothesis, the strength of associations of most symptoms was not significantly different from those of mood or emotional involvement. The two exceptions were early reduction in sleep and the appetite, whose associations with overall depression severity at post-treatment were significantly weaker than those of mood and emotional involvement. Instead, for seven of the nine symptoms, associations were of similar magnitude with standardized betas around 0.4 (mood, feelings of unease, ability to concentrate, initiative, emotional involvement, pessimism, and zest for life), while sleep and appetite had standardized betas of 0.27. Translated into raw MADRS-S points, each one-step reduction in a specific symptom until week four was associated with roughly three points lower MADRS-S score at post-treatment for seven of the symptoms, while for sleep and appetite each one-step was associated with roughly two points lower MADRS-S score at post-treatment. Results were similar in the cross-lagged network analysis, where early change in most symptoms showed similar associations with symptom-specific severity at post-treatment. Again, the two exceptions were sleep and appetite, for which early change were less associated with level of other symptoms at post-treatment. These results are in line with prior network analyses findings that symptoms of sleep and appetite are relatively weakly related to other depression symptoms (Boschloo et al., 2016; Fried et al., 2016; Hakulinen et al., 2020).

If taken as causal estimates, our results indicate that – except for sleep and appetite – early reductions in most specific depression symptoms have similar effects on later depression levels. Interpreting any of these associations as causal effects rests on several formal assumptions that are discussed below. Informally, it seems plausible that change in one symptom could have effects on other symptoms at later time-points. Theories guiding psychological treatments frequently assume causal interactions between symptoms, although details of these relationships are often vaguely described (Haslbeck et al., 2022). For example, that reducing pessimistic thinking could lead to improved mood, as proposed in Beck's cognitive model of depression (Wright & Beck, 1983) or that increased initiative could lead to improved mood and emotional involvement by positive reinforcement, as proposed in the behavioural activation model of depression (Jacobson et al., 2001). From a treatment perspective, our results, with its similar associations across symptoms, suggests limited benefits of targeting any specific depression symptom over another early in treatment, at least at the group level. On a conceptual level, our results

dovetails meta-analytic evidence that different psychological treatments, targeting different mechanisms and symptoms, have similar effects on depression (Cuijpers et al., 2021) with limited evidence for the effects of specific treatment components (Cuijpers et al., 2019). Importantly, our results, as well as those of most treatment studies, are averaged at the group level, and for individual patients there may well be benefits of targeting some specific symptoms over others. It has been suggested that the heterogeneity of the depression diagnosis may in part explain the failure to single out specific mechanisms of change at the group level (Maj et al., 2020). It is therefore possible that the similarity of estimates across most of the symptoms examined could in part be explained by diagnostic heterogeneity, and that more refined characterizations of patients may be needed to clearly differentiate out influential symptoms (cf. Bringmann, 2024).

Sleep and appetite deviated from the above-described pattern. It should be noted that these symptoms were also not targeted early in the ICBT-treatment which could potentially have influenced the results (appetite was not targeted in the ICBT program at all, and sleep problems were covered in the eight module). Decreased appetite was the symptom for which early change were least associated with subsequent depression. This was also the symptom that was least endorsed at pre-treatment, as measured by its mean value. Together, these findings suggest that appetite problems may be a relatively marginal problem in those seeking help trough ICBT, compared to other depression symptoms, and that improved appetite may not impact overall depression symptomatology to any large extent. Sensitivity analyses showed stronger associations between early change in appetite symptom and later depression among individuals with worse initial depression (although individuals with severe MDD were not included in the cohort). This could suggest that appetite is of greater importance when the depression level is higher, but may also reflect a greater variance in appetite in the high depression groups (Lee et al., 2023; Terluin et al., 2016). Our results also suggest that improvements in sleep may be of less importance in relation to overall depression severity, versus improvements in the majority of other specific symptoms of depression. This is somewhat at odds with prior research, where targeting sleep problems using ICBT for insomnia has shown promising results in reducing depression among individuals with depression and comorbid insomnia even after three years (Blom et al., 2015, 2017). However, another study with this comorbid patient group indicates that combining ICBT for depression with ICBT for insomnia did not improve depression more than ICBT for depression plus insomnia placebo (Blom et al., 2024) Disturbed sleep is arguably one of the least characteristic symptoms of depression, as it is frequently present across many different mental disorders (Baglioni et al., 2016). It is possible that sleep may be of high importance in subpopulations of depressed patients, such as those with comorbid insomnia, but less influential than more specific depression symptoms in the general depressed population.

It should be noted that the similar associations across symptoms are also compatible with a latent construct view on depression. From this perspective symptoms are seen as interchangeable indicators of an underlying depressive condition, rather than causally efficacious treatment targets themselves (as in the network theory). When symptoms are assumed to be indicators of a latent construct, their associations to an outcome are expected to be proportional to their reliability as measures of the latent construct (e.g. their factor loadings) (VanderWeele, 2022; VanderWeele & Vansteelandt, 2022). Sleep and appetite symptoms have been found to be less reliable indicators of depression in some factor analytic studies of the MADRS-S (Ntini et al., 2020; Quilty et al., 2013), which could explain their weaker associations to later overall depression severity in our analyses.

Strengths and limitations

There are several strengths of this study. First, unlike most other network studies on interactions between specific depression symptoms (McNally, 2021), we used a longitudinal design allowing for stronger evidence of causality. This allowed for control for prior symptom levels along with a variety of other potentially confounding variables which strengthens the plausibility of the no residual and unmeasured confounding assumption required for causal interpretations of estimates (see Supplemental Methods 3 for an elaboration), although this assumption can never be proven with observational data. Second, although there was some loss to follow-up, the use of multiple imputation to handle missing data lessens the risk for selection bias. Third, causal interpretation also relies on the consistency assumption (VanderWeele, 2022), that is, that the exposures (i.e. symptoms) are sufficiently well-defined such that there are no multiple versions of the same symptom that have different effects on later depression. Symptoms may hold different meanings, for instance the mood symptom may reflect different kinds of depressed mood, but our focus on individual symptoms of depression increases the granularity and specificity of our results. Fourth, we conducted several sensitivity analyses showing that associations cannot be explained solely by autoregressive effects, the choice of depression scale or the way that change was modelled. It has been suggested that differences in connection strength can be explained by differences in the variance of the included variables (Lee et al., 2023; Terluin et al., 2016). This explanation seems implausible in the present study, given that two symptoms (emotional involvement and zest for life) had lower variance than appetite and sleep was the symptom with the second highest variance at week four (Supplementary Figure 4).

The study also has limitations. First, the time-scale for measuring symptom change may not be optimal. Change was measured during the first four weeks, a decision based on prior research showing that this is a suitable time-window to predict post-treatment outcomes (Forsell et al., 2020; Schibbye et al., 2014). It is possible, however, that more fine-grained time-scales (i.e. weeks, days, or hours) may be needed to accurately capture interactions between symptoms. Symptoms interacting on shorter time-scales may have acted as time-varying confounders (and mediators) in our analyses leading to overlap and diluted differences between our estimates, analogous to when multiple interacting mediators are analysed in separate models (cf. VanderWeele, 2015). Second, no individuals with severe MDD were included in the cohort, which is of importance as our sensitivity analyses showed stronger associations among individuals with higher initial depression. It is therefore possible that our results do not generalize to patients with more severe depression. Third, the use of single items to measure symptoms increases the risk of measurement error affecting the results, which is a limitation shared with much of the current psychological network research (de Ron et al., 2022). However, sensitivity analyses showed that similar results were obtained when we used repeated symptom-measures as predictors in the form of individual regression slopes, that should be more robust to random measurement error.

Conclusions

We found that early reduction in all nine depression symptoms measured by the MADRS-S were associated with lower depression severity at post-treatment, both at the overall and at the symptom-specific levels. Contrary to our hypothesis, early change in the hallmark depression symptoms (mood and emotional involvement) did not stand out as more strongly associated with post-treatment depression severity compared with most other symptoms. Instead, associations were of similar magnitude for seven of the symptoms, with only sleep and appetite showing weaker associations with

depression severity at post-treatment. Interpreting our results as causal effects rests on several assumptions, but our findings suggest that the impact of early reduction in specific symptoms on depression severity at post-treatment in ICBT for MDD may be similar across most symptoms, but less for the sleep and appetite symptoms.

Acknowledgements

We would like to thank the patients and the clinicians at the Internet Psychiatry Unit in Region Stockholm for their valuable contribution to this research.

Financial support

This research was supported by the Swedish Research Council (2021-06377 JW; 2018-02487 CR; 2016-01961 VK), the Swedish Research Council for Health, Working Life and Welfare (2018-00221 and 2021-00132 CR), The Söderström König Foundation (SLS- 941192 and SLS-994792 JW) and The Center for Innovative Medicine – CIMED (96328 JW).

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.

Competing interests

The authors declare no competing interests.

References

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders: DSM-5*. American Psychiatric Association.

Baglioni, C., Nanovska, S., Regen, W., Spiegelhalder, K., Feige, B., Nissen, C., Reynolds, C. F., & Riemann, D. (2016). Sleep and mental disorders: A meta-analysis of polysomnographic research. *Psychol Bull*, *142*(9), 969–990. https://doi.org/10.1037/bul0000053

Beard, J. I. L., & Delgadillo, J. (2019). Early response to psychological therapy as a predictor of depression and anxiety treatment outcomes: A systematic review and meta-analysis. *Depression and Anxiety*, *36*(9), 866–878. https://doi.org/10.1002/da.22931

Blom, K., Forsell, E., Hellberg, M., Svanborg, C., Jernelöv, S., & Kaldo, V. (2024). Psychological Treatment of Comorbid Insomnia and Depression: A Double-Blind Randomized Placebo-Controlled Trial. *Psychotherapy and Psychosomatics*, *93*(2), 100–113. https://doi.org/10.1159/000536063

Blom, K., Jernelöv, S., Kraepelien, M., Bergdahl, M. O., Jungmarker, K., Ankartjärn, L., Lindefors, N., & Kaldo, V. (2015). Internet treatment addressing either insomnia or depression, for patients with both diagnoses: A randomized trial. *Sleep*, *38*(2), 267–277. https://doi.org/10.5665/sleep.4412

Blom, K., Jernelöv, S., Rück, C., Lindefors, N., & Kaldo, V. (2017). Three-Year Follow-Up Comparing Cognitive Behavioral Therapy for Depression to Cognitive Behavioral Therapy for Insomnia, for Patients With Both Diagnoses. *Sleep*, *40*(8), zsx108. https://doi.org/10.1093/sleep/zsx108

Boberg, J., Kaldo, V., Mataix-Cols, D., Crowley, J. J., Roelstraete, B., Halvorsen, M., Forsell, E., Isacsson, N. H., Sullivan, P. F., Svanborg, C., Andersson, E. H., Lindefors, N., Kravchenko, O., Mattheisen, M., Danielsdottir, H. B., Ivanova, E., Boman, M., Cruz, L. F. de la, Wallert, J., & Rück, C. (2023). Swedish multimodal cohort of patients with anxiety or depression treated with internet-delivered psychotherapy (MULTI-PSYCH). *BMJ Open*, *13*(10), e069427. https://doi.org/10.1136/bmjopen-2022-069427

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*(1), 91–121. https://doi.org/10.1146/annurev-clinpsy-050212-185608

Boschloo, L., van Borkulo, C. D., Borsboom, D., & Schoevers, R. A. (2016). A Prospective Study on How Symptoms in a Network Predict the Onset of Depression. *Psychother Psychosom*, *85*(3), 183–184. https://doi.org/10.1159/000442001

Bringmann, L. F. (2024). The future of dynamic networks in research and clinical practice. *World Psychiatry*, *23*(2), 288–289. https://doi.org/10.1002/wps.21209

Cuijpers, P., Cristea, I. A., Karyotaki, E., Reijnders, M., & Hollon, S. D. (2019). Component studies of psychological treatments of adult depression: A systematic review and meta-analysis. *Psychotherapy Research*, *29*(1), 15–29. https://doi.org/10.1080/10503307.2017.1395922

Cuijpers, P., Quero, S., Noma, H., Ciharova, M., Miguel, C., Karyotaki, E., Cipriani, A., Cristea, I. A., & Furukawa, T. A. (2021). Psychotherapies for depression: A network meta-analysis covering efficacy, acceptability and long-term outcomes of all main treatment types. *World Psychiatry*, *20*(2), 283–293. https://doi.org/10.1002/wps.20860

de Ron, J., Robinaugh, D. J., Fried, E. I., Pedrelli, P., Jain, F. A., Mischoulon, D., & Epskamp, S. (2022). Quantifying and addressing the impact of measurement error in network models. *Behaviour Research and Therapy*, *157*, 104163. https://doi.org/10.1016/j.brat.2022.104163

Forsell, E., Isacsson, N., Blom, K., Jernelöv, S., Ben Abdesslem, F., Lindefors, N., Boman, M., & Kaldo, V. (2020). Predicting treatment failure in regular care Internet-Delivered Cognitive Behavior Therapy for depression and anxiety using only weekly symptom measures. *Journal of Consulting and Clinical Psychology*, *88*(4), 311–321. https://doi.org/10.1037/ccp0000462

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. https://doi.org/10.1016/j.jad.2015.09.005

Fried, E. I., & Nesse, R. M. (2015). Depression is not a consistent syndrome: An investigation of unique symptom patterns in the STAR*D study. *Journal of Affective Disorders*, *172*, 96–102. https://doi.org/10.1016/j.jad.2014.10.010

Garke, M. Å., Hentati Isacsson, N., & Kaldo, V. (2021). *A Possibly Mean Mean: The Importance of Considering Heterogenous Change in Discrete Symptoms when Defining Outcomes of Cognitive Behavioral Therapy for Patients with Anxiety and Mood Disorders*. https://doi.org/10.31234/osf.io/q6532

Hakulinen, C., Fried, E. I., Pulkki-Råback, L., Virtanen, M., Suvisaari, J., & Elovainio, M. (2020). Network structure of depression symptomology in participants with and without depressive disorder: The population-based Health 2000–2011 study. *Social Psychiatry and Psychiatric Epidemiology*, *55*(10), 1273–1282. https://doi.org/10.1007/s00127-020-01843-7

Haslbeck, J. M. B., Ryan, O., Robinaugh, D. J., Waldorp, L. J., & Borsboom, D. (2022). Modeling psychopathology: From data models to formal theories. *Psychological Methods*, *27*(6), 930–957. https://doi.org/10.1037/met0000303

Hedman, E., Ljótsson, B., Kaldo, V., Hesser, H., El Alaoui, S., Kraepelien, M., Andersson, E., Rück, C., Svanborg, C., Andersson, G., & Lindefors, N. (2014). Effectiveness of Internet-based cognitive behaviour therapy for depression in routine psychiatric care. *Journal of Affective Disorders*, *155*, 49–58. https://doi.org/10.1016/j.jad.2013.10.023

Jacobson, N. S., Martell, C. R., & Dimidjian, S. (2001). Behavioral Activation Treatment for Depression: Returning to Contextual Roots. *Clinical Psychology: Science and Practice*, *8*(3), 255–270. https://doi.org/10.1093/clipsy.8.3.255

Kopta, S. M., Howard, K. I., Lowry, J. L., & Beutler, L. E. (1994). Patterns of Symptomatic Recovery in Psychotherapy. *Journal of Consulting and Clinical Psychology*, *62*(5), 1009–1016. https://doi.org/10.1037/0022-006X.62.5.1009

Kroenke, K., Spitzer, R. L., & Williams, J. B. (2001). The PHQ-9: Validity of a brief depression severity measure. *J Gen Intern Med*, *16*(9), 606–613. https://doi.org/10.1046/j.1525-1497.2001.016009606.x

Lee, C. T., Kelley, S. W., Palacios, J., Richards, D., & Gillan, C. M. (2023). *Estimating theprognostic value of cross-sectional network connectivity for treatment response indepression*. OSF. https://doi.org/10.31234/osf.io/agfc8

Maj, M., Stein, D. J., Parker, G., Zimmerman, M., Fava, G. A., De Hert, M., Demyttenaere, K., McIntyre, R. S., Widiger, T., & 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

McNally, R. J. (2021). Network Analysis of Psychopathology: Controversies and Challenges. *Annu Rev Clin Psychol*, *17*, 31–53. https://doi.org/10.1146/annurev-clinpsy-081219-092850

Nakamura, J. S., Hong, J. H., Smith, J., Chopik, W. J., Chen, Y., VanderWeele, T. J., & Kim, E. S. (2022). Associations Between Satisfaction With Aging and Health and Well-being Outcomes Among Older US Adults. *JAMA Network Open*, *5*(2), e2147797–e2147797. https://doi.org/10.1001/jamanetworkopen.2021.47797

Ntini, I., Vadlin, S., Olofsdotter, S., Ramklint, M., Nilsson, K. W., Engström, I., & Sonnby, K. (2020). The Montgomery and Åsberg Depression Rating Scale – self-assessment for use in adolescents: An evaluation of psychometric and diagnostic accuracy. *Nordic Journal of Psychiatry*, *74*(6), 415–422. https://doi.org/10.1080/08039488.2020.1733077

Paternoster, R., Brame, R., Mazerolle, P., & Piquero, A. (1998). Using the Correct Statistical Test for the Equality of Regression Coefficients. *Criminology*, *36*(4), 859–866. https://doi.org/10.1111/j.1745-9125.1998.tb01268.x

Quilty, L. C., Robinson, J. J., Rolland, J.-P., Fruyt, F. D., Rouillon, F., & Bagby, R. M. (2013). The structure of the Montgomery-Åsberg depression rating scale over the course of treatment for depression. *International Journal of Methods in Psychiatric Research*, *22*(3), 175–184. https://doi.org/10.1002/mpr.1388

Robinaugh, D. J., Millner, A. J., & McNally, R. J. (2016). Identifying highly influential nodes in the complicated grief network. *J Abnorm Psychol*, *125*(6), 747–757. https://doi.org/10.1037/abn0000181

Schibbye, P., Ghaderi, A., Ljótsson, B., Hedman, E., Lindefors, N., Rück, C., & Kaldo, V. (2014). Using Early Change to Predict Outcome in Cognitive Behaviour Therapy: Exploring Timeframe, Calculation Method, and Differences of Disorder-Specific versus General Measures. *PLOS ONE*, *9*(6), e100614. https://doi.org/10.1371/journal.pone.0100614

Solmi, M., Cortese, S., Vita, G., De Prisco, M., Radua, J., Dragioti, E., Köhler-Forsberg, O., Madsen, N. M., Rohde, C., Eudave, L., Aymerich, C., Pedruzo, B., Rodriguez, V., Rosson, S., Sabé, M., Hojlund, M., Catalan, A., de Luca, B., Fornaro, M., … Correll, C. U. (2023). An umbrella review

of candidate predictors of response, remission, recovery, and relapse across mental disorders. *Molecular Psychiatry*. https://doi.org/10.1038/s41380-023-02298-3

Svanborg, P., & Åsberg, M. (1994). A new self-rating scale for depression and anxiety states based on the Comprehensive Psychopathological Rating Scale. *Acta Psychiatrica Scandinavica*, *89*(1), 21–28. https://doi.org/10.1111/j.1600-0447.1994.tb01480.x

Szegedi, A., Jansen, W. T., van Willigenburg, A. P., van der Meulen, E., Stassen, H. H., & Thase, M. E. (2009). Early improvement in the first 2 weeks as a predictor of treatment outcome in patients with major depressive disorder: A meta-analysis including 6562 patients. *J Clin Psychiatry*, *70*(3), 344–353. https://doi.org/10.4088/jcp.07m03780

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

van Ballegooijen, W., Riper, H., Cuijpers, P., van Oppen, P., & Smit, J. H. (2016). Validation of online psychometric instruments for common mental health disorders: A systematic review. *BMC Psychiatry*, *16*(1), 45. https://doi.org/10.1186/s12888-016-0735-7

VanderWeele, T. J. (2015). *Explanation in Causal Inference: Methods for Mediation and Interaction*. Oxford University Press.

VanderWeele, T. J. (2022). Constructed Measures and Causal Inference: Towards a New Model of Measurement for Psychosocial Constructs. *Epidemiology*, *33*(1), 141–151. https://doi.org/10.1097/ede.0000000000001434

VanderWeele, T. J., Jackson, J. W., & Li, S. (2016). Causal inference and longitudinal data: A case study of religion and mental health. *Soc Psychiatry Psychiatr Epidemiol*, *51*(11), 1457–1466. https://doi.org/10.1007/s00127-016-1281-9

VanderWeele, T. J., & Vansteelandt, S. (2022). A statistical test to reject the structural interpretation of a latent factor model. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *84*(5), 2032–2054. https://doi.org/10.1111/rssb.12555

Wichers, M., Riese, H., Hodges, T. M., Snippe, E., & Bos, F. M. (2021). A Narrative Review of Network Studies in Depression: What Different Methodological Approaches Tell Us About Depression. *Frontiers in Psychiatry*, *12*. https://doi.org/10.3389/fpsyt.2021.719490

World Health Organization. (n.d.). *International statistical classification of diseases and related health problems* (11th ed., Vol. 2019).

Wright, J. H., & Beck, A. T. (1983). Cognitive Therapy of Depression: Theory and Practice. *Psychiatric Services*, *34*(12), 1119–1127. https://doi.org/10.1176/ps.34.12.1119

Wysocki, A., Rhemtulla, M., Van Bork, R., & Cramer, A. (2022). Cross-lagged Network Models. *PsyArXiv*. https://doi.org/10.31234/osf.io/vjr8z