Psychological Medicine

cambridge.org/psm

Depression networks: a systematic review of the network paradigm causal assumptions

Debbie Huang[1,2], Ezra Susser[2,3], Kara E. Rudolph[2] and Katherine M. Keyes[2]

[1]Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA, USA; [2]Department of Epidemiology, Mailman School of Public Health, Columbia University, New York, NY, USA and [3]New York State Psychiatric Institute, New York, New York, United States of America

Review Article

Cite this article: Huang D, Susser E, Rudolph KE, Keyes KM (2023). Depression networks: a systematic review of the network paradigm causal assumptions. *Psychological Medicine* **53**, 1665–1680. https://doi.org/10.1017/S0033291723000132

Received: 20 April 2022
Revised: 7 December 2022
Accepted: 12 January 2023
First published online: 17 March 2023

Keywords:
Causality; depression; measurement; network paradigm; psychopathology

Author for correspondence:
Debbie Huang,
E-mail: debbiehuang@hsph.harvard.edu

Abstract

The network paradigm for psychiatric disorder nosology was proposed based on the hypothesis that mental disorders are caused by networks of symptoms that are themselves causally related. Researchers have widely applied and integrated this paradigm to examine a variety of mental disorders, particularly depression. Existing studies generally focus on the correlation structure of symptoms, inferring causal relationships. Thus, presumption of causality may not be justified. The goal of this review was to examine the assumptions necessary for causal inference in network studies of depression. Specifically, we examined whether and how network studies address common violations of causal assumptions (i.e. no measurement error, exchangeability, and positivity). Of the 41 studies reviewed, five (12%) studies discussed sources of confounding unrelated to measurement error; none discussed positivity; and five conducted post-hoc analysis for measurement error. Depression network studies, in principle, are conducted under the assumption that symptom relationships are causal. Yet, in practice, studies seldomly discussed or adequately tested assumptions required to infer causality. Researchers continue to design studies that are unable to support the credibility of the network paradigm for the study of depression. There is a critical need to ensure scientific efforts cease to perpetuate problematic designs and findings to a potentially unsubstantiated paradigm.

Background

The network paradigm was proposed based on the thought-provoking hypothesis that mental disorders are networks made up of mental health symptoms causing each other (Borsboom, 2008; Fried, 2015; Kendler, Zachar, & Craver, 2011). Under the network paradigm, 'disorders' are complex, mutually reinforcing systems of symptoms that are causally related to each other (Fried, 2015; Kendler et al., 2011). The paradigm is an extension of earlier theoretical work (Cramer et al., 2016; Hayes & Strauss, 1998; Kendler et al., 2011; Schiepek & Tschacher, 1992; Teasdale, 1983) which conceptualized mental disorders as emergent properties and dynamic complex system. The fundamental assumption of this paradigm is that symptom relationships are directly causally related to each other, rather than being correlated due to arising from the same underlying construct. For example, a network hypothesis may be that depressed mood caused self-blame. Self-blame then caused insomnia. As a result of being fatigued, an inability to sleep is exacerbated and an individual has problems concentrating. The disorder of depression is made up of this network of causal chains. There may be multiple different symptom pathways that can be activated, resulting in heterogeneity in the presentation of the disorder. There may be a lag in effect on the system, and the system may be self-sustained via feedback loops (Borsboom, 2017; Cramer et al., 2016).

This fundamental assumption of symptoms causing each other contrasts with the traditional common cause paradigm. In the traditional paradigm, the assumption is that symptom relationships are correlational and that those correlations arise because symptoms are caused by an underlying disorder that is not directly measurable (American Psychiatric Association, 2013; Nesse & Stein, 2012). For example, common symptoms of depression include depressed mood, self-blame, insomnia, fatigue, and concentration problems. Under the common cause paradigm, these symptoms are associated with each other because the underlying dimension of depression caused each symptom. There has been great interest in contrasting the two paradigms but is predicated on the validity of the network paradigm.

The adoption of the network paradigm has gained popularity over the past decade, particularly for the study of depression. Since 2008, there have been more than 240 network studies on mental health psychopathology. During this time, researchers have widely applied and integrated this paradigm to examine a variety of mental health topics including, but not limited to, comorbidity, remission and onset of symptoms, development psychopathology, biological psychiatry, etc. Nearly one-third of empirical network studies have examined depressive

© The Author(s), 2023. Published by Cambridge University Press

symptoms (Robinaugh, Hoekstra, Toner, & Borsboom, 2020). Studies that have adopted the network paradigm have yielded two consistent findings regarding the psychopathology of depression. First, a recent systematic review of 65 empirical network studies characteristics (e.g. sample type, instrument estimated, and reliability of findings) (Contreras, Nieto, Valiente, Espinosa, & Vazquez, 2019) found a multitude of depressive symptom associations. Moreover, these associations remained even after controlling for other depressive symptoms and demographic covariates (e.g. age, gender, and depression severity status). Second, studies suggest symptoms appear to mutually reinforce each other over time. Intraindividual analyses of depression experience revealed interaction and lags between different mood states and symptoms that are consistent with the network paradigm (aan het Rot, Hogenelst, & Schoevers, 2012; Bringmann et al., 2013; van Borkulo et al., 2015; Wichers, 2014). By examining symptom dynamics, network studies have begun to uncover some of the complex mechanisms of depression, making the network paradigm potentially a useful tool for furthering our understanding of depression. However, the reviews also identified several major limitations with depression network studies.

Although the examination of the network theory may be valuable for the study of mental disorders, researchers have also raised concerns about the value and credibility of existing depression network studies. To date, three systematic network literature reviews (Contreras et al., 2019; Malgaroli, Calderon, & Bonanno, 2021; Robinaugh et al., 2020), two narrative reviews (Guloksuz, Pries, & van Os, 2017; Wichers, Riese, Hodges, Snippe, & Bos, 2021), and a number of commentary papers (Borsboom et al., 2021; Fried & Cramer, 2017) have noted limitations of the network literature. First, there is a reliance on cross-sectional data and the production of undirected correlation networks, which are not appropriate for causal inference. Undirected graphs do not contain information about directionality and consist of joint correlations between nodes that may or may not account for the fact that correlations may be due to shared relationships with other symptoms. There are very few studies that produce directed networks, which utilize longitudinal data and depict the temporal ordering between symptoms. Furthermore, inferring causation from observation data requires additional assumptions (e.g. exchangeability and positivity). Without the verification of such assumptions, network findings can only be, at best, interpreted as predictive over time (i.e. Granger's causality) and cannot be interpreted as causal (Granger, 1969). The lack of empirical evidence for symptom causal relationships is a critical threat to the paradigm and studies which adopt this paradigm.

Second, researchers remain apprehensive about the accuracy of mental health measures. Network studies' results lack reliability or fail to converge across study. Studies often focus on identifying the most influential symptom as measured by centrality indices, which are statistical measures that summarize and identify important node relationships within a network. However, studies have failed to converge on a set of central symptoms. In a recent review of 23 cross-sectional depression network studies, the most commonly reported central symptoms were depressed mood (10/32 networks) and fatigue (9/32 networks) but each of these were central symptoms in less than one-third of the studies (Malgaroli et al., 2021). In addition to the lack of convergence on a central symptom, studies have also failed to converge on symptom patterns. For instance, anhedonia (i.e. the experience of loss of pleasure or interest) has been found to be a central symptom in some studies (Boschloo, van Borkulo, Borsboom, & Schoevers, 2016; Bringmann, Lemmens, Huibers, Borsboom, & Tuerlinckx, 2015; Fried, Epskamp, Nesse, Tuerlinckx, & Borsboom, 2016) but found to be the most peripheral symptom (Berlim, Richard-Devantoy, Dos Santos, & Turecki, 2020; Kendler, Aggen, Flint, Borsboom, & Fried, 2018) in other studies. It is unclear whether the lack of convergence in findings is due to chance or methodological error. However, the lack of replicability hinders generalizability and identification of meaningful difference between group and time. The lack of convergent findings has been the subject of debates (Borsboom et al., 2017; Forbes, Wright, Markon, & Krueger, 2017a, 2017b, 2021; Funkhouser et al., 2020; Jones, Williams, & McNally; Steinley, Hoffman, Brusco, & Sher, 2017). Measurement error in the assessment of depressive symptoms is a potential explanation for the inconsistency within depression network studies and needs to be further investigated.

There are several limitations with the current network paradigm review studies. First, beyond critiques about the use of cross-sectional datasets, reviews have not fully assessed whether the existing studies met the assumptions for supporting causal relationships among symptoms. In a seminal paper, Borsboom (2017) defined a causal relationship between symptoms as, 'the presence of a causal connection means that, if an (experimental or natural) intervention changed the state of one symptom, this would change the probability distribution of the other symptom'. This definition reflects a counterfactual approach to causal inference. Positivity and exchangeability are the major assumptions for identifying the causal effect (Pearl, 2000; Schwartz, Gatto, & Campbell, 2016). Exchangeability is the lack of confounding, due to common causes of the exposure and outcome of interest (Greenland & Robins, 2009; VanderWeele, 2019). Positivity assumes there is an individual for every observed combination of exposure and covariates in the contrast of interest, which prevents off-support inference (Petersen, Porter, Gruber, Wang, & van der Laan, 2012). In a recent narrative review (Wichers et al., 2021), the authors discussed concerns about the possible influence of third variables for the network connections but did not elaborate on the issue beyond importance of node selection. Other existing reviews also have not fully investigated nor discussed the underlying causal framework and assumptions.

Second, despite the concerns about the convergence of results and accuracy of mental health measures, the reviews do not elaborate on the extent to which measurement error has been investigated. In the context of depression networks, measurement error is the inability to capture all or aspects of depressive symptoms or other variables that are to be included in the network model of interest (Buonaccorsi, 2010). Measurement of depressive symptoms, in general, is subject to numerous sources of error – participants may forget their symptoms, misunderstand the question, and/or feel embarrassed and misreport their symptoms. Further, previous depression network studies have relied on scales that measure a varying number of depressive symptoms, at various time periods, and with heterogeneous underlying validity for measuring depression (Malgaroli et al., 2021). Since each symptom could be susceptible to measurement error, the cumulative effect of measurement error across symptoms could result in different network structures and centrality indices across studies. Measurement error is a common source of unobserved non-exchangeability (Hernán & Robins, 2020; Pearl, 2000) and it is also related to violations of consistency, another causal assumption, which assumes the exposure is sufficiently measured so that there are no variations in exposure which could influence

https://doi.org/10.1017/S0033291723000132 Published online by Cambridge University Press

the outcome (Cole & Frangakis, 2009; Rehkopf, Glymour, & Osypuk, 2016; VanderWeele, 2009). The consistency assumptions have been the subject of numerous debates. Many social exposure and network approaches violate the interference criterion (Bhattacharya, Malinsky, & Shpitser, 2019; Schwartz, Gatto, & Campbell, 2012; VanderWeele & An, 2013) for this assumption. As such, for this review, we will focus on the impact of measurement error rather than an evaluation of the consistency assumption.

An in-depth investigation of these issues within the network literature is needed prior to the widespread application of this paradigm. A systematic review investigating depression network studies on the credibility of their assumptions would close the existing knowledge gap. Thus, the aim of the review is to synthesize the existing literature to (1) evaluate the reliability of network findings by examining whether centrality results are converging or not and (2) evaluate the quality of empirical studies by assessing whether studies are designed to support symptom causal relationships. The second aim will also include an evaluation of the quality of empirical studies by assessing whether they have examined the impact of measurement error.

Methods

Registration

The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement was followed for the conduct of the systematic review. The methods of the search, search terms, inclusion/exclusion criteria, and analysis were registered in advance with PROSPERO.

Search strategy

The systematic review sought to capture all English, full-text articles of depression network studies, published between 2008 and 2021. Since the network paradigm was first proposed in 2008, the search excluded studies prior to this year. The search was conducted in the PubMed, ISI, Web of Science, and PsycINFO databases, which are all major databases in capturing various psychiatry, psychology, data science, and public health journals. Search terms were broad and included: ('network approach' OR 'network analysis' OR 'network perspective' OR 'causal system' OR 'symptom network') AND ('depression' or 'depressive'). Reference lists of the identified articles were further screened to capture any articles missed by the search terms or databases.

Screening and selection: inclusion and exclusion criteria

Search results were imported into Covidence for systematic review management. After removing the duplicate studies, titles and abstracts were screened to ensure the search criteria (i.e. English, peer-reviewed, full-text, and between 2008 and 2021) have been met.

There were two steps in the screening process. In the first step, two reviewers (DH and SM) independently reviewed titles and abstracts. Proportion of agreement was 83%. Disagreements (n = 156) were then discussed until reaching consensus. The selected full-text articles were then assessed according to the following inclusion and exclusion criteria. (1) Studies were related to the network paradigm. Previous reviews had not provided explicit criteria for what constituted a network study (Contreras et al., 2019; Guloksuz et al., 2017; Malgaroli et al., 2021; Robinaugh et al., 2020; Wichers et al., 2021). However, for the present review, a depression network study was defined as a study that had the results which produced (a) a network graph or (b) centrality results of depressive symptoms. These screening criteria excluded any study, for example, which evaluated social networks, brain circuitry networks, and healthcare system networks. (2) Studies included measures of depressive symptoms. Unlike previous reviews, which examined a variety of psychological phenomenon, the current review only focused on depression. Studies where the unit of analysis was another psychiatric disorder (e.g. dysthymia, generalized anxiety disorder, post-traumatic stress disorder), were excluded. (3) Studies examined only depression networks. Network studies that examined the network of depression with another disorder or discussed 'bridge symptoms' in the findings were excluded.

In the second step of the screening process, full articles were then categorized by the study type – theoretical (n = 8), methodological (n = 9), or empirical (n = 41). Theoretical articles were studies that provided an explanation for the network paradigm or theory but did not provide any empirical analysis. Methodological articles were studies that reported empirical results, but the goal was to demonstrate the robustness of the network model to a specific bias or test models for a specific assumption. Findings in methodological studies inform how statistical analysis should be conducted. Empirical articles were studies that reported analytic results, and the findings informed the psychopathology of depressive symptoms. The purpose of this review was to evaluate the quality of studies that adopted the network paradigm to understand the psychopathology of depressive symptoms. Thus, only empirical studies were included in this review. A summary diagram of the search and screening is shown in Fig. 1.

Data extraction and analysis

The current review is unique from other reviews (Contreras et al., 2019; Malgaroli et al., 2021; Robinaugh et al., 2020) in that it explores the causal assumptions within the network literature. As such, for this analysis, we consulted risk of bias tools and other causal inference measures to select a list of indicators that would be relevant for network studies and our study goals. The following pieces of information were extracted from the articles: source population, study design (i.e. cross-sectional, longitudinal), sample size, network analytic design (i.e. un-directed network, directed network), the depressive instrument used, number of symptoms measured, symptom connectivity, comparison groups (i.e. age, gender, severity group), and symptom centrality.

Symptom centrality is generally measured by centrality indices. Common indices for undirected networks included strength, betweenness, and closeness. Strength described how strongly nodes are connected to each other. Closeness provided information on the distance from a node to other nodes in the system. Betweenness described how often a node acted as a mediator or bridge between two other nodes, which provided information on how strongly a node can disrupt information flow within a network (Dalege, Borsboom, van Harreveld, & van der Maas, 2017). Common centrality indices within directed network included in-strength/degree and out-strength/degree. In-strength referred to reactivity to other symptoms. Out-strength referred to the likelihood of impacting other symptoms. Nodes with high values on the common indices are the most central on that index. When more than one index was used in a study, the

https://doi.org/10.1017/S0033291723000132 Published online by Cambridge University Press

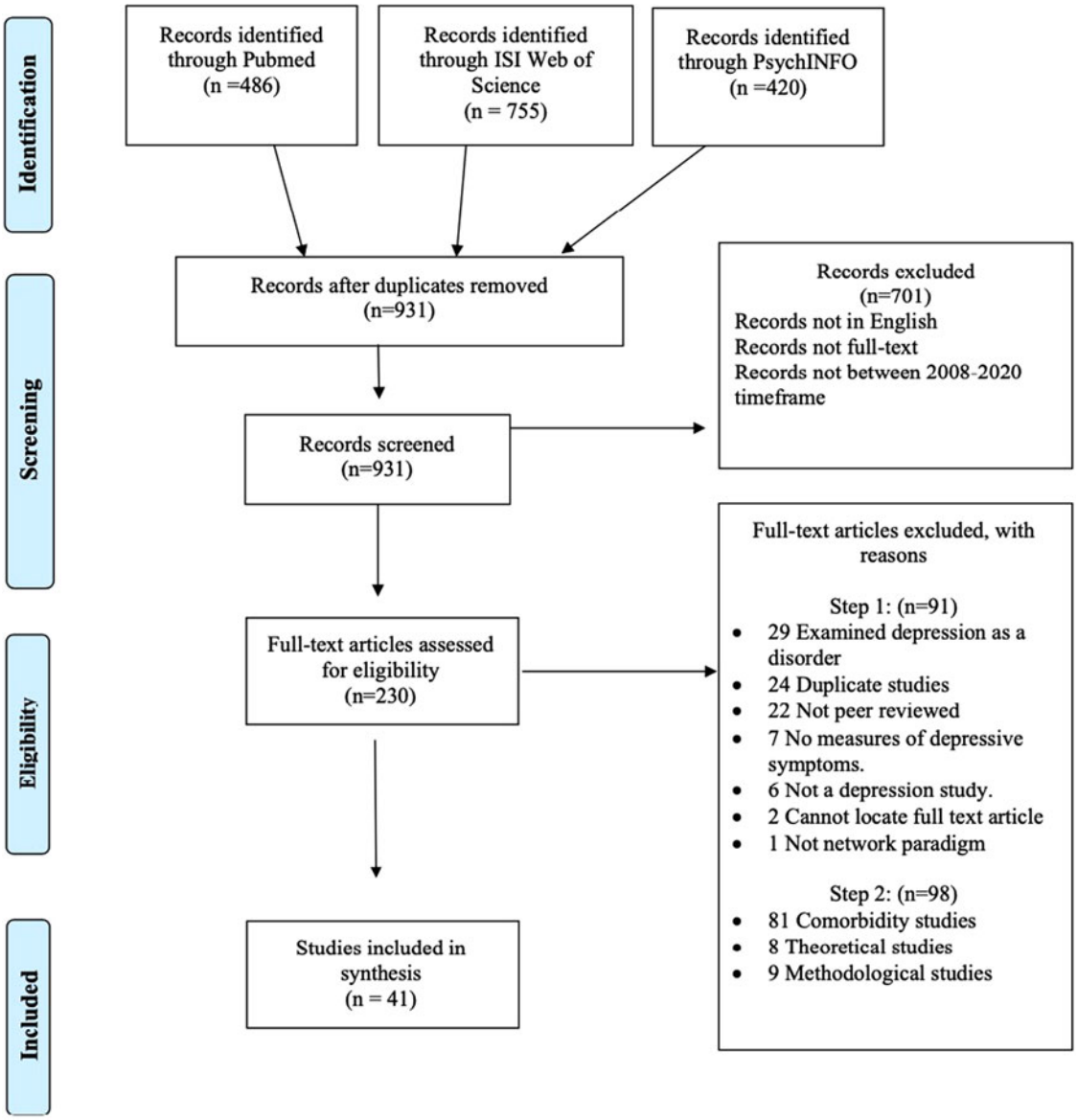


Fig. 1. PRISMA flow diagram – depression network.

most central node is generally determined by the judgment of the investigator. While there have been arguments that betweenness and closeness may be unsuitable measures for the psychological networks (Bringmann et al., 2019; Hallquist, Wright, & Molenaar, 2021), in this review, central symptoms reported by any of these indices were extracted. For example, if the study found the symptoms sadness had a high betweenness score and fatigue had a strength score, the review would record both sadness and fatigue as central symptoms.

The studies were evaluated based on the study design, sample size, analytic design, adjustments for non-exchangeability, and assessment or discussion of other causal assumptions. Study design, sample size, and analytic design are important considerations in existing study quality assessment tools such as GRADE and STROBE. Any adjustments or discussion of two causal assumptions (i.e. exchangeability or positivity), assessment of positivity, addressing non-exchangeability (e.g. regression adjustment, control group), whether studies made any causal claims about depressive symptoms within its conclusion, and discussion of causality within its limitation were recorded (see Supplementary 1). A paper is considered to have made a causal conclusion only if the authors had explicitly stated the results could be interpreted as causal. Any discussions of instrument validity and reliability (i.e. psychometrics of instrument), the impact of measurement error (i.e. amount, source), and any post-hoc or sensitivity analysis to evaluate the impact of measurement error were also documented and included as considerations for non-exchangeability.

Results

Study characteristics

There were 41 empirical network studies identified which focused on depressive symptoms. The majority (63.41%) of these studies were published after 2019. The sample size of studies ranged

https://doi.org/10.1017/S0033291723000132 Published online by Cambridge University Press

https://doi.org/10.1017/S0033291723000132 Published online by Cambridge University Press

Table 1. Depression network studies characteristics

Type	Authors (year)	Population	Sample size	Instrument	Number of symptoms	Central symptoms	Discussed causal assumptions (i.e. exchangeability or positivity)	Assessed positivity	Adjustments for non-exchangeability	Made causal claims in conclusion	Mentioned causality in limitation	Discussed instrument validity (i.e. validity, reliability)	Discussed measurement error (i.e. amount, type)	Addressed measurement error
Undirected cross-sectional depression studies characteristics ($N = 25$)	Cramer et al. (2012)	Community	2096	Disaggregated symptoms of major depression	14	Worthless-ness and suicide ideation	No	No	No	No	Yes	No	No	No
	Song et al. (2015)	Clinical	364	Traditional Chinese medicine depression symptom	36	NA	No	No	No	No	No	No	No	No
	Fried et al. (2016)	Clinical	3463	IDS	30	Sadness and fatigue	No	No	Yes – controlled for other symptoms in the network	No	Yes	Yes	No	Yes – collapsed weight problem and appetite problem so only 28 total symptoms
	McWilliams et al. (2017)	Clinical	216	PHQ-9	9	Sadness, loss of interest/pleasure, fatigue, and concentration problems	No	No	Yes – controlled for other symptoms in the network	No	Yes	Yes – mentioned scale items are instrument dependent	No	No
	Santos et al. (2017)	Community	515	CES-D; Becks	41	Sadness	No	No	Yes – controlled for other symptoms in the network	No	No	Yes – reliability and the fact that measures might be overlapping	No	No
	Kendler et al. (2018)	Community	5952	CIDI and other non-DSM criteria	19	Psychomotor change	No	No	Yes – controlled for other symptoms in the network	No	Yes	No	No	No
	van Loo et al. (2018)	Community	5784	CIDI	24	NA	No	No	Yes – controlled for other symptoms in the network	No	Yes	Yes – mentioned scale items are instrument dependent	Yes (self-report)	Yes – sensitivity analysis of recall bias
	Hartung et al. (2019)	Clinical	8040	PHQ-9	9	Sadness, fatigue, worthlessness, and suicide ideation	No	No	Yes – controlled for other symptoms in the network	No	Yes	Yes	No	No
	Mullarkey et al. (2019)	Community	1409	Children's depression inventory	27	Sadness and worthlessness	No	No	Yes – controlled for other symptoms in the network	No	No	Yes	No	No
	Baez and Heller (2020)	Clinical	3184	HDRS	17	Sadness, worthlessness, and sleep problems	Yes	No	Yes – age, length of current episode, number of past episode, and quality of life	No	Yes	No	Yes (self-report)	No

(*Continued*)

https://doi.org/10.1017/S0033291723000132 Published online by Cambridge University Press

Table 1. (*Continued.*)

Type	Authors (year)	Population	Sample size	Instrument	Number of symptoms	Central symptoms	Discussed causal assumptions (i.e. exchangeability or positivity)	Assessed positivity	Adjustments for non-exchangeability	Made causal claims in conclusion	Mentioned causality in limitation	Discussed instrument validity (i.e. validity, reliability)	Discussed measurement error (i.e. amount, type)	Addressed measurement error
	Belvederi Murri et al. (2020a)	Community	8557	EURO_D	12	Sadness, loss of interest/pleasure, and suicide ideation	No	No	No	No	Yes	No	No	No
	Belvederi Murri et al. (2020b)	Clinical	447	PHQ-9	9	Sadness, worthlessness, and suicide	No	No	No	No	Yes	No	No	No
	Briganti et al. (2020)	Community	1090	Self-rating depression scale	20	NA	Yes	No	Yes – controlled for other symptoms in the network	Yes – causation can be inferred from DAG results assuming no confounding or sampling bias	No	No	No	No
	Burger et al. (2020)	Community	724	CES-D	15	NA	No	No	Yes – controlled for other symptoms in the network, marital status	No	Yes	Yes	No	Yes – collapsed items that were correlated $\geqslant .5$ and ended up with 12 items rather than 15
	Castellanos et al. (2020)	Community	555	CIDI	9	Sadness, loss of interest/pleasure, and sleep problems	No	No	Yes – controlled for other symptoms in the network	No	No	Yes	No	No
	Corponi et al. (2020)	Clinical	2758	DSM-IV-TR	23	Psychomotor agitation	No	No	Yes – controlled for other symptoms in the network	No	Yes	No	No	No
	de la Torre-Luque et al. (2020)	Clinical	427	CIDI	12	Psychomotor agitation and concentration problems	No	No	Yes – controlled for other symptoms in the network	No	No	Yes – mentioned scale items are instrument dependent	No	No
	de Vos et al. (2020)	Community	254 443	Kessler psychological distress scale	10	Sadness, psychomotor agitation, and worthlessness	No	No	Yes – controlled for other symptoms in the network	No	Yes	No	No	Yes – collapsed redundant variables so only 50 total symptoms

https://doi.org/10.1017/S0033291723000132 Published online by Cambridge University Press

	Fried et al. (2020)	Clinical	2321	IDS	30	NA	No	No	Yes – controlled for other symptoms in the network, age, sex, alcohol, smoking, BMI, number of chronic disease being treated, and physical activity	No	Yes	Yes – mentioned scale items are instrument dependent	No	No
	Hakulinen et al. (2020)	Community	6593	Becks	13	Sadness, loss of interest/pleasure, fatigue, and worthlessness	No	No	Yes – control of without diagnosis	No	Yes	No	Yes (important symptoms may be missing from using BDI)	No
	Lass et al. (2020)	Clinical	1042	Becks; distress tolerance scale; and behavior activation for depression scale	60	Suicide ideation	No	No	Yes – controlled for other symptoms in the network	No	Yes	Yes – reliability and the fact that measures might be overlapping	Yes (self-report)	Yes – collapsed redundant variables so only 50 total symptoms
	Park et al. (2020)	Clinical	1174	International classification of diseases-10	10	Sadness, loss of interest/pleasure, and fatigue	No	No	Yes – controlled for other symptoms in the network	No	No	No	No	No
	Gijzen et al. (2021)	Community	5888	Children's depression inventory	28	Sadness and worthlessness	No	No	Yes – controlled for other symptoms in the network	No	Yes	Yes	No	No
	Pan and Liu (2021)	Community	484	CES-D	20	Sadness	Yes	No	Yes – controlled for other symptoms in the network, age, gender, education, religious faith, hukou, and income. Also had controls	No	Yes	Yes – mentioned scale items are instrument dependent	No	No
	Vetter et al. (2021)	Clinical	590	HDRS	17	Sadness and psychomotor agitation			Yes – controlled for other symptoms in the network and severity	No	No	Yes – reliability and the fact that measures might be overlapping	No	No
Undirected longitudinal depression studies characteristics ($N = 11$)	Fried et al. (2015)	Community	515	CES-D	11	NA	No	No	Yes – controlled for other symptoms in the network	No	No	Yes – mentioned scale items are instrument dependent	No	No
	Koenders et al. (2015)	Clinical	125	Young mania rating scale and quick IDS	27	Sadness and psychomotor agitation	No	No	No	No	Yes	Yes	No	No
	Madhoo et al. (2016)	Clinical	2876	Quick IDS-self report	14	Sadness and fatigue	No	No	Yes – controlled for other symptoms in the network	No	No	No	No	No

(*Continued*)

https://doi.org/10.1017/S0033291723000132 Published online by Cambridge University Press

Table 1. (*Continued.*)

Type	Authors (year)	Population	Sample size	Instrument	Number of symptoms	Central symptoms	Discussed causal assumptions (i.e. exchangeability or positivity)	Assessed positivity	Adjustments for non-exchangeability	Made causal claims in conclusion	Mentioned causality in limitation	Discussed instrument validity (i.e. validity, reliability)	Discussed measurement error (i.e. amount, type)	Addressed measurement error
	van Borkulo et al. (2016)	Clinical	515	IDS	9	Decreased interest and fatigue	Yes	No	Yes – controlled for other symptoms in the network and severity	No	No	No	No	No
	Semino et al. (2017)	Clinical	110	Becks	21	Emotional regulation	No	No	Yes – controlled for other symptoms in the network	No	No	Yes	No	No
	Bos et al. (2018)	Clinical	178	Becks	178	Loss of interest and fatigue	No	No	Yes – RCT healthy controls	No	Yes	No	No	No
	McElroy et al. (2019)	Clinical	3017	Revised Children's Anxiety and Depression Scale	10	Sadness and fatigue	No	No	Yes – controlled for other symptoms in the network and matched on baseline depressive severity score	No	No	Yes – mentioned scale items are instrument dependent	No	No
	Airaksinen et al. (2020)	Community	7779	CES-D	8	Sadness	No	No	Yes – participants matched on age, sex, and ethnic background	No	No	No	Yes (self-report)	No
	Berlim et al. (2020)	Clinical	151	Quick IDS-self report	9	Sadness and fatigue	No	No	Yes – RCT healthy controls	No	Yes	Yes	Yes (self-report)	No
	Mullarkey et al. (2020)	Community	295	Quick IDS-self report	16	NA	No	No	Yes – controlled for other symptoms in the network	No	No	Yes – reliability and the fact that measures might be overlapping	Yes (self-report)	No
	Smetter et al. (2021)	Community	177	Becks (extended)	22	Loss of interest/ pleasure and fatigue	No	No	Yes – controlled for other symptoms in the network	No	Yes	Yes	No	No
Directed network depression studies characteristics ($N = 5$)	Dejonckheere et al. (2017)	Community	112	Symptoms based on DSM-5	11	Core symptom – in-strength anhedonia; cognitive depression – out-strength social expectancy; positive somatic symptom – out-strength social expectancies and psychomotor high in-strength; negative somatic symptom fatigue had high out-strength and psychomotor high in-strength	No	No	Yes – controlled for other symptoms in the network	No	Yes	No	No	No

https://doi.org/10.1017/S0033291723000132 Published online by Cambridge University Press

	Yang et al. (2018)	Community	150	CES-D	20	NA	No	No	Yes – life events	No	No	Yes – mentioned that the field is moving toward dimensionality	No	No
	Aalbers et al. (2019)	Community	125	Symptoms based on DSM	7	Fatigue and loneliness	No	No	Yes – controlled for other symptoms in the network, stress, passive social media use, and active social media use	No	No	No	No	No
	Groen et al. (2019)	Clinical	69	Symptom checklist-90-revised and HDRS	30	For persistent group – 'feeling everything is an effort had highest positive out-strength'; for reduced group – 'worrying too much about things appear to have highest positive out-strength'	Yes	No	Yes – severity	No	No	No	No	No
	Savelieva et al. (2021)	Community	72 971	EURO_D	12	Sadness and diminished interest and suicide ideation	No	No	Yes – age, sex, mean level of depression, and follow-up time	No	No	Yes	Yes (self-report)	No

PHQ-9, Patient Health Questionnaire-9; HDRS, Hamilton Rating Scale for Depression; Becks, Beck Depression Inventory-II; IDS, Inventory of Depressive Symptomatology; CES-D, Center of Epidemiological Studies Depression Scale; ICD, International Classification of Diseases; CIDI, Composite International Diagnostic Interview; DSM, Diagnostic and Statistical Manual of Mental Disorders.

https://doi.org/10.1017/S0033291723000132 Published online by Cambridge University Press

Table 2. Depression network studies characteristics by measures

Authors (year)	Population	Sample size	Instrument	Type	Number of symptoms	Central symptoms
Hakulinen et al. (2020)	Community	6593	Becks	Undirected cross-sectional	13	Sadness, loss of interest/pleasure, fatigue, and worthlessness
Semino et al. (2017)	Clinical	110	Becks	Undirected longitudinal	21	Emotional regulation
Bos et al. (2018)	Clinical	178	Becks	Undirected longitudinal	178	Loss of interest and fatigue
Smetter et al. (2021)	Community	177	Becks (extended)	Undirected longitudinal	22	Loss of interest/pleasure and fatigue
Lass et al. (2020)	Clinical	1042	Becks; distress tolerance scale; and behavior activation for depression scale	Undirected cross-sectional	60	Suicide ideation
Santos et al. (2017)	Community	515	CES-D; Becks	Undirected cross-sectional	41	Sadness
Burger et al. (2020)	Community	724	CES-D	Undirected cross-sectional	15	NA
Pan and Liu (2021)	Community	484	CES-D	Undirected cross-sectional	20	Sadness
Fried et al. (2015)	Community	515	CES-D	Undirected longitudinal	11	NA
Airaksinen et al. (2020)	Community	7779	CES-D	Undirected longitudinal	8	Sadness
Yang et al. (2018)	Community	150	CES-D	Directed longitudinal	20	NA
Mullarkey et al. (2019)	Community	1409	Children's depression inventory	Undirected cross-sectional	27	Sadness and worthlessness
Gijzen et al. (2021)	Community	5888	Children's depression inventory	Undirected cross-sectional	28	Sadness and worthlessness
van Loo et al. (2018)	Community	5784	CIDI	Undirected cross-sectional	24	NA
de la Torre-Luque et al. (2020)	Clinical	427	CIDI	Undirected cross-sectional	12	Psychomotor agitation and concentration problems
Castellanos et al. (2020)	Community	555	CIDI	Undirected cross-sectional	9	Sadness, loss of interest/pleasure, and sleep problems
Kendler et al. (2018)	Community	5952	CIDI and other non-DSM criteria	Undirected cross-sectional	19	Psychomotor change
Belvederi Murri et al. (2020a)	Community	8557	EURO_D	Undirected cross-sectional	12	Sadness, loss of interest/pleasure, and suicide ideation
Savelieva et al. (2021)	Community	72 971	EURO_D	Directed longitudinal	12	Sadness and diminished interest and suicide ideation

https://doi.org/10.1017/S0033291723000132 Published online by Cambridge University Press

Baez and Heller (2020)	Clinical	3184	HDRS	Undirected cross-sectional	17	Sadness, worthlessness, and sleep problems
Vetter et al. (2021)	Clinical	590	HDRS	Undirected cross-sectional	17	Sadness and psychomotor agitation
Groen et al. (2019)	Clinical	69	Symptom checklist-90-revised and HDRS	Directed longitudinal	30	For persistent group – 'feeling everything is an effort had highest positive out-strength'; for reduced group – 'worrying too much about things appear to have highest positive out-strength'
Fried et al. (2016)	Clinical	3463	IDS	Undirected cross-sectional	30	Sadness and fatigue
Fried et al. (2020)	Clinical	2321	IDS	Undirected cross-sectional	30	NA
van Borkulo et al. (2016)	Clinical	515	IDS	Undirected longitudinal	9	Decreased interestand fatigue
Park et al. (2020)	Clinical	1174	International classification of diseases-10	Undirected cross-sectional	10	Sadness, loss of interest/pleasure, and fatigue
de Vos et al. (2020)	Community	254 443	Kessler psychological distress scale	Undirected cross-sectional	10	Sadness, psychomotor agitation, and worthlessness
McWilliams et al. (2017)	Clinical	216	PHQ-9	Undirected cross-sectional	9	Sadness**,** loss of interest/pleasure, fatigue, and concentration problems
Hartung et al. (2019)	Clinical	8040	PHQ-9	Undirected cross-sectional	9	Sadness, fatigue, worthlessness, and suicide ideation
Belvederi Murri et al. (2020b)	Clinical	447	PHQ-9	Undirected cross-sectional	9	Sadness, worthlessness, and suicide
Madhoo et al. (2016)	Clinical	2876	Quick IDS-self report	Undirected longitudinal	14	Sadness and fatigue
Berlim et al. (2020)	Clinical	151	Quick IDS-self report	Undirected longitudinal	9	Sadness and fatigue
Mullarkey et al. (2020)	Community	295	Quick IDS-self report	Undirected longitudinal	16	NA
Koenders et al. (2015)	Clinical	125	Young mania rating scale and quick IDS	Undirected longitudinal	27	Sadness and psychomotor agitation
McElroy et al. (2019)	Clinical	3017	Revised children's anxiety and depression scale	Undirected longitudinal	10	Sadness and fatigue
Mullarkey et al. (2019)	Community	1409	Children's depression inventory	Undirected cross-sectional	27	Sadness and worthlessness
Gijzen et al. (2021)	Community	5888	Children's depression inventory	Undirected cross-sectional	28	Sadness and worthlessness
Briganti et al. (2020)	Community	1090	Self-rating depression scale	Undirected cross-sectional	20	NA
Aalbers et al. (2019)	Community	125	Symptoms based on DSM	Directed longitudinal	7	Fatigue and loneliness

(*Continued*)

Table 2. (*Continued.*)

Authors (year)	Population	Sample size	Instrument	Type	Number of symptoms	Central symptoms
Dejonckheere et al. (2017)	Community	112	Symptoms based on DSM-V	Directed longitudinal	11	Core symptom – in-strength anhedonia; cognitive depression – out-strength social expectancy; positive somatic symptom – out-strength social expectancies and psychomotor high in-strength; negative somatic symptom fatigue had high out-strength and psychomotor high in-strength
Cramer et al. (2012)	Community	2096	Disaggregated symptoms of major depression	Undirected cross-sectional	14	Worthless-ness and suicide ideation
Corponi et al. (2020)	Clinical	2758	DSM-IV-TR	Undirected cross-sectional	23	Psychomotor agitation
Song et al. (2015)	Clinical	364	Traditional Chinese medicine Depression symptom	Undirected cross-sectional	36	NA

PHQ-9, Patient Health Questionnaire-9; HDRS, Hamilton Rating Scale for Depression; Becks, Beck Depression Inventory-II; IDS, Inventory of Depressive Symptomatology; CES-D, Center of Epidemiological Studies Depression Scale; ICD, International Classification of Diseases; CIDI, Composite International Diagnostic Interview; DSM, Diagnostic and Statistical Manual of Mental Disorders.

from 69 to 254 443, with a median of 724. Half of the studies were conducted among community samples ($n = 21$, median sample size = 1090), and half were conducted among clinical samples ($n = 20$, median sample size = 553). There were 36 undirected network studies: 25 cross-sectional studies, and 11 longitudinal studies (all panel sampling). There were five studies (one panel and four intensive sampling) that reported directed network results (see Table 1 and Supplementary 2). Depressive symptoms were most assessed using the Beck's Depression Inventory-II and Center of Epidemiological Studies Depression Scale. The median number of symptoms measured was 16 (range of 7–178 symptoms).

Convergence of centrality findings by study type

Centrality findings varied by the type of network study – cross-sectional undirected, longitudinal undirected, and directed. Within undirected cross-sectional network studies, the centrality results were highly variable and inconsistent. That is, findings across studies did not converge to suggest one symptom as the most central. The most reported central symptoms in undirected cross-sectional network studies were sadness and worthlessness. There were 15 studies that reported sadness as one of the central symptoms (15 highest strength, 4 betweenness, and 3 closeness). There were eight studies that reported worthlessness as one of the central symptoms (seven highest strength, three betweenness, and two closeness). Within undirected longitudinal studies, fatigue was the most reported central symptom at baseline networks and sadness was the most central symptom at follow-up networks. Within directed network studies, centrality results also greatly varied by study.

Convergence of centrality findings by instrument

There was less heterogeneity in centrality results when examined within studies that used similar instruments (see Table 2). For studies which had utilized the Beck's Depression Inventory-II, three of the six studies had reported loss of interest and fatigue as the most central symptoms. Sadness was reported as the most central symptom in studies which used the Center of Epidemiological Studies Depression Scale (three out of six studies), Patient Health Questionnaire-9 (all three studies), and Inventory of Depressive Symptomatology (three out of four studies).

Casual assumptions and measurement error

Of the 41 studies assessed, 70.3% had some discussion related to an assumption of causal inference (see Table 3). But only five of the studies had mentioned sources of confounding or non-exchangeability that were not related to measurement error. None of the studies assessed positivity. About 88% adjusted for other depressive symptoms as potential confounders of the depressive symptom relationships. This included the five studies which had conducted sensitivity or post-hoc analysis to examine the impact of measurement error. For 46.3% of the studies, the inability to infer causation was mentioned as a limitation.

Discussion of causal assumptions was slightly between cross-sectional and longitudinal studies (see Tables 4 and 5). Sixty-eight percent of cross-sectional studies and 75% of longitudinal studies had some discussion of causal assumptions. A greater proportion of longitudinal studies (25% *v.* 16%) had

https://doi.org/10.1017/S0033291723000132 Published online by Cambridge University Press

Table 3. All depression network study causal inference assumptions ($n = 41$)

Characteristic	Number of studies	Percentage of studies
Sample size (⩾1000)	20	48.78
Network type		
Undirected	25	60.98
Undirected longitudinal	11	26.83
Directed	5	12.20
Discussed casual assumption	29	70.73
Discussed positivity	0	0.00
Discussed sources of exchangeability	5	12.20
Provided psychometrics of instruments	23	56.*10*
Discussed source of measurement error	8	19.51
Assessed for positivity	0	0.00
Adjusted for exchangeability	36	87.80
Adjusted for measurement error with post-hoc/sensitivity analysis	5	12.20
Made a causal conclusion	1	2.44
Mentioned causality within limitation	19	46.34

Table 4. Cross-sectional studies depression network study causal inference assumptions ($n = 25$)

Characteristic	Number of studies	Percentage of studies
Sample size (⩾1000)	16	64.00
Network type		
Undirected	25	100
Undirected longitudinal	0	0.00
Directed	0	0.00
Discussed casual assumption	17	68.00
Discussed positivity	0	0.00
Discussed sources of exchangeability	3	12.00
Provided psychometrics of instruments	14	56.00
Discussed source of measurement error	4	16.00
Assessed for positivity	0	0.00
Adjusted for exchangeability	21	84.00
Adjusted for measurement error with post-hoc/sensitivity analysis	5	20.00
Made a causal conclusion	1	4.00
Mentioned causality within limitation	8	32.00

discussed sources of measurement error. However, none of the longitudinal studies had conducted sensitivity or post-hoc analysis to examine the impact of measurement error. About 69% of longitudinal studies had mentioned causality within its limitation, whereas only 32% of cross-sectional had discussed this topic.

Discussion

Depression network studies have expanded, particularly in the past 2 years. This expansion is a demonstration of the increasing interest to adopt the network approaches to study depression. However, the review revealed several challenges that continue to impede the credibility of network studies and the network paradigm.

First, most depression network studies are not capable of providing empirical support for symptom causal relationships. Studies are conducted under the assumption that symptom relationships are causal, but the study design elides establishing causal relationships from correlations. Symptom correlations arise due to confounding, such as by a shared latent construct. Results may be explained by the network paradigm, the common cause paradigm, or unmeasured factors. Since many depression network studies are still mostly conducted with cross-sectional data or are analyzed cross-sectionally, producing undirected network results, these studies are unable to demonstrate the precedence of symptoms and do not meet the assumptions for identifying causal effects. The reliance on cross-sectional data has been criticized by previous reviews of the general network literature (Contreras et al., 2019; Guloksuz et al., 2017; Malgaroli et al., 2021; Robinaugh et al., 2020). Results from the current review further found that this practice continues within newer depression network studies.

Depression network studies had rarely assessed the full series of causal assumptions. In this review, approximately 88% of the studies adjusted for sources of non-exchangeability. The majority of the adjustment had been limited to only existing study symptoms within the network. However, to isolate the independent association between two symptoms, adjustment of another symptom is not sufficient. Previous studies have found factors such as age (Colman & Ataullahjan, 2010), gender (Piccinelli & Wilkinson, 2000), and severity (Kessler, Chiu, Demler, Merikangas, & Walters, 2005) are important contributors to the depression experience. Consideration of these factors and other potential sources of common causes of depressive symptoms should also be considered in future network studies.

Also, positivity was never assessed in any of the studies. In general, sparse or no data may suggest violation of positivity. Violation of positivity is important as estimates may be biased if one group has a zero probability, or close to zero probability, of experiencing a symptom conditional on observed covariates. For example, a comparison of individuals that experienced sadness and did not experience sadness is not valid or logical if data from one of the groups were not collected. As previously stated, most studies adjust for other existing symptoms within the network. The tendency to adjust for many covariates within an analysis suggests positivity is likely to be violated. Thus, depression network studies, in principle, are conducted under the assumption that symptom causal relationships exist. However, in practice, studies seldomly discussed or adequately tested these criteria.

Depression network studies must be designed to better inform a potential causal relationship between depressive symptoms. Future studies should attempt to (1) utilize longitudinal data

https://doi.org/10.1017/S0033291723000132 Published online by Cambridge University Press

Table 5. Longitudinal studies depression network study causal inference assumptions ($n = 16$)

Characteristic	Number of studies	Percentage of studies
Sample size (⩾1000)	4	25.00
Network type		
Undirected	0	0.00
Undirected longitudinal	11	68.75
Directed	5	31.25
Discussed casual assumption	12	75.00
Discussed positivity	0	0.00
Discussed sources of exchangeability	2	12.50
Provided psychometrics of instruments	9	56.25
Discussed source of measurement error	4	25.00
Assessed for positivity	0	0.00
Adjusted for exchangeability	15	93.75
Adjusted for measurement error with post-hoc/sensitivity analysis	0	0.00
Made a causal conclusion	0	0.00
Mentioned causality within limitation	11	68.75

and analytic designs, and (2) assess causal inference assumptions. This includes considering all potential sources of non-exchangeability and incorporate more rigorous methods (e.g. inverse probability weighting) for adjusting for non-exchangeability. Also, studies must take caution to not violate positivity. Studies incorporate both longitudinal study design and assessment of causal inference assumptions would better support potential causal relationships and demonstrate the credibility of the network paradigm.

Second, depression network studies neglected to investigate the impact of measurement error. In the context of depression network studies, measurement error is the inability to capture all or aspects of depressive symptoms (Buonaccorsi, 2010). However, other than recall bias, most studies do not discuss the validity and reliability of the instrument used or explore the impact measurement error. This is particularly concerning given the fact that this review also found that centrality results appeared to vary across different studies, and instability remains a critical concern (Forbes, Wright, Markon, & Krueger, 2019; Forbes et al., 2017a, 2021; Funkhouser et al., 2020; Guloksuz et al., 2017; Malgaroli et al., 2021; Wichers et al., 2021). Sadness was generally the most reported central symptom. This was consistent with previous reviews (Contreras et al., 2019; Guloksuz et al., 2017; Malgaroli et al., 2021; Wichers et al., 2021). But there was a great deal of heterogeneity, depending on the study type and instrument used. Considering that depressive symptoms were also measured by a variety of instruments; measurement error may be an important contributor to the difference in findings.

Measurement of depressive symptoms is subject to numerous sources of error. Underreporting of symptoms and recall bias are common sources of measurement error (Wells & Horwood, 2004). Different instruments assessed different numbers of symptoms, and symptoms are assessed with different levels of precision. For example, the Becks Depression Inventory (BDI) assessed 21 depressive symptoms (Beck, Ward, Mendelson, Mock, & Erbaugh, 1961). The Patient Health Questionnaire (PHQ-9) only assessed nine symptoms (Kroenke, Spitzer, & Williams, 2001). And while the BDI assessed loss of interest and loss of pleasure as separate symptoms, the PHQ-9 assessed for both symptoms together as anhedonia. Due to these varying sources of error and precision, at any given moment, each symptom could be impacted by measurement error.

Measurement error may be ever-present and depending on the amount and type of measurement error, it can impact the connection between symptoms. Measurement error is a common form of unobserved source of non-exchangeability. The cumulative effect of measurement error across symptoms could result in different network structures and centrality indices. Results from previous robustness studies of non-mental health network studies have found centrality indices are prone to be influenced by measurement error (Borgatti, Carley, & Krackhardt, 2006; Frantz, Cataldo, & Carley, 2009; Kim & Jeong, 2007; Martin & Niemeyer, 2019). The robustness of mental health network findings to measurement error should be investigated.

In consideration of the issue of reliability and measurement error, future investigations should consider the following recommendations. First, studies should, at a minimum, discuss the reliability and validity of instruments. Studies need to better consider the number of symptoms measured, how well depressive symptoms are captured, that symptoms reported typically change over (even short) time intervals and discuss or explore the impact of measurement error on the interpretation of results. Also, studies are needed to examine the robustness of depression networks to measurement error. This would be useful for understanding how measurement error may contribute to inconsistent results across depression network studies.

Limitations

The current study has the following limitations. First, the search was conducted using three major databases (PubMed, ISI Web of Science, and PsycINFO databases), which should be sufficient in capturing various psychiatry, psychology, data science, and public health journals. Reference lists of the identified articles were also screened to capture any missed articles. However, it is still possible that studies published in smaller, international, or specialty journals were missed.

Second, the review excluded 81 depression comorbidity studies. There is a large body of literature that examined the network structure of depressive symptoms and other mental disorders. These studies focused on identifying symptoms that link disorders together (i.e. 'bridge symptoms'). Consideration of causality and measurement error are more complex in these studies. There are potentially more sources of non-exchangeability and measurement error since symptoms from multiple disorders must be considered. Examination of the comorbidity network studies between depression and other mental disorders should be the focus of future investigations.

Third, the proposed study utilizes self-made indices to evaluate how well current studies evaluate and support the assumptions of the depression network. Study characteristics were chosen based on existing evaluation tools, causal frameworks, and previous studies. However, these indices merely act as a tool to summarize

https://doi.org/10.1017/S0033291723000132 Published online by Cambridge University Press

the existing network findings and are not validated measures of study causal validity. Future studies could consider different or more in-depth criteria.

Conclusion

The goal of the study is to evaluate depression network studies on their discussion and evaluation of the network paradigm's causal assumptions. There is growing enthusiasm to adopt the network approach. However, there are important gaps within this literature that must be addressed. Future network researchers should take caution when designing studies. There is a need to empirically test the causal relationship among symptoms and test the robustness of centrality results to measurement error. Prior to the widespread utilization of these techniques, the fundamental assumptions of this paradigm must be explored.

Supplementary material. The supplementary material for this article can be found at https://doi.org/10.1017/S0033291723000132.

Acknowledgment. The authors thank Dr Sarah McKetta for her assistance with the screening process.

Financial support. This work was supported by the National Institute of Mental Health (DH, grant numbers NIH T32 MH013043 and T32 MH 017119).

Conflict of interest. The authors have no conflict of interest to declare.

References

aan het Rot, M., Hogenelst, K., & Schoevers, R. A. (2012). Mood disorders in everyday life: A systematic review of experience sampling and ecological momentary assessment studies. *Clinical Psychology Review*, *32*(6), 510–523. doi: 10.1016/j.cpr.2012.05.007

American Psychiatric Association, DSM-5 Task Force. (2013). *Diagnostic and statistical manual of mental disorders: DSM-5™* (5th ed.). Arlington, VA: American Psychiatric Publishing, Inc. doi: 10.1176/appi.books.9780890425596.

Beck, A. T., Ward, C. H., Mendelson, M., Mock, J., & Erbaugh, J. (1961). An inventory for measuring depression. *Archives of General Psychiatry*, *4*(6), 561–571. doi: 10.1001/archpsyc.1961.01710120031004

Berlim, M. T., Richard-Devantoy, S., Dos Santos, N. R., & Turecki, G. (2020). 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. doi: 10.1017/s0033291720001002.

Bhattacharya, R., Malinsky, D., & Shpitser, I. (2019). *Causal inference under interference and network uncertainty. Uncertainty in artificial intelligence: Proceedings of the Conference on Uncertainty in Artificial Intelligence, 2019.*

Borgatti, S. P., Carley, K. M., & Krackhardt, D. (2006). On the robustness of centrality measures under conditions of imperfect data. *Social Networks*, *28*(2), 124–136. doi: 10.1016/j.socnet.2005.05.001

Borsboom, D. (2008). Psychometric perspectives on diagnostic systems. *Journal of Clinical Psychology*, *64*(9), 1089–1108. doi:https://doi.org/10.1002/jclp.20503

Borsboom, D. (2017). A network theory of mental disorders. *World Psychiatry*, *16*(1), 5–13. doi: 10.1002/wps.20375

Borsboom, D., Deserno, M. K., Rhemtulla, M., Epskamp, S., Fried, E. I., McNally, R. J., … Waldorp, L. J. (2021). Network analysis of multivariate data in psychological science. *Nature Reviews Methods Primers*, *1*(1), 58. doi: 10.1038/s43586-021-00055-w

Borsboom, D., Fried, E. I., Epskamp, S., Waldorp, L. J., van Borkulo, C. D., van der Maas, H. L. J., & Cramer, A. O. J. (2017). False alarm? A comprehensive reanalysis of 'Evidence that psychopathology symptom networks have limited replicability' by Forbes, Wright, Markon, and Krueger (2017). *Journal of Abnormal Psychology*, *126*(7), 989–999. doi: 10.1037/abn0000306

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. *Psychotherapy and Psychosomatics*, *85*(3), 183–184. doi: 10.1159/000442001

Bringmann, L. F., Elmer, T., Epskamp, S., Krause, R. W., Schoch, D., Wichers, M., … Snippe, E. (2019). What do centrality measures measure in psychological networks? *Journal of Abnormal Psychology*, *128*(8), 892–903. doi: 10.1037/abn0000446

Bringmann, L. F., Lemmens, L., Huibers, M. J. H., Borsboom, D., & Tuerlinckx, F. (2015). Revealing the dynamic network structure of the Beck Depression Inventory-II. *Psychological Medicine*, *45*(4), 747–757. doi: 10.1017/s0033291714001809

Bringmann, L. F., Vissers, N., Wichers, M., Geschwind, N., Kuppens, P., Peeters, F., … Tuerlinckx, F. (2013). A network approach to psychopathology: New insights into clinical longitudinal data. *PLoS ONE*, *8*(4), e60188. doi: 10.1371/journal.pone.0060188

Buonaccorsi, J. (2010). *Measurement error*. New York, NY: Chapman and Hall/CRC.

Cole, S. R., & Frangakis, C. E. (2009). The consistency statement in causal inference: A definition or an assumption? *Epidemiology (Cambridge, Mass.)*, *20*(1), 3–5. doi: 10.1097/EDE.0b013e31818ef366

Colman, I., & Ataullahjan, A. (2010). Life course perspectives on the epidemiology of depression. *The Canadian Journal of Psychiatry. Revue canadienne de psychiatrie*, *55*(10), 622–632. doi: 10.1177/070674371005501002

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. doi: 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. doi: 10.1371/journal.pone.0167490

Dalege, J., Borsboom, D., van Harreveld, F., & van der Maas, H. L. J. (2017). Network analysis on attitudes: A brief tutorial. *Social Psychological and Personality Science*, *8*(5), 528–537. doi: 10.1177/1948550617709827

Forbes, M. K., Wright, A. G. C., Markon, K. E., & Krueger, R. F. (2019). Quantifying the reliability and replicability of psychopathology network characteristics. *Multivariate Behavioral Research*, *56(2)*, 224–242. doi: 10.1080/00273171.2019.1616526.

Forbes, M. K., Wright, A. G. C., Markon, K. E., & Krueger, R. F. (2017a). Evidence that psychopathology symptom networks have limited replicability. *Journal of Abnormal Psychology*, *126*(7), 969–988. doi: 10.1037/abn0000276

Forbes, M. K., Wright, A. G. C., Markon, K. E., Krueger, R. F. (2017b). Further evidence that psychopathology networks have limited replicability and utility: Response to Borsboom *et al.* (2017) and Steinley *et al.* (2017). *Journal of Abnormal Psychology*, *126*(7), 1011–1016. doi: 10.1037/abn0000313

Forbes, M. K., Wright, A. G. C., Markon, K. E., & Krueger, R. F. (2021). On unreplicable inferences in psychopathology symptom networks and the importance of unreliable parameter estimates. *Multivariate Behavioral Research*, *56*(2), 368–376. doi: 10.1080/00273171.2021.1886897

Frantz, T. L., Cataldo, M., & Carley, K. M. (2009). Robustness of centrality measures under uncertainty: Examining the role of network topology. *Computational and Mathematical Organization Theory*, *15*(4), 303. doi: 10.1007/s10588-009-9063-5

Fried, E. I. (2015). Problematic assumptions have slowed down depression research: Why symptoms, not syndromes are the way forward. *Frontiers in Psychology*, *6*, 309–309. doi: 10.3389/fpsyg.2015.00309

Fried, E. I., & Cramer, A. O. J. (2017). Moving forward: Challenges and directions for psychopathological network theory and methodology. *Perspectives on Psychological Science*, *12*(6), 999–1020. doi: 10.1177/1745691617705892

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. *Journal of Affective Disorders*, *189*, 314–320. doi: 10.1016/j.jad.2015.09.005

Funkhouser, C. J., Correa, K. A., Gorka, S. M., Nelson, B. D., Phan, K. L., & Shankman, S. A. (2020). The replicability and generalizability of internalizing symptom networks across five samples. *Journal of Abnormal Psychology*, *129*(2), 191–203. doi: 10.1037/abn0000496

Granger, C. W. J. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, *37*(3), 424–438. doi: 10.2307/1912791

Greenland, S., & Robins, J. M. (2009). Identifiability, exchangeability and confounding revisited. *Epidemiologic Perspectives & Innovations*, *6*(1), 4. doi: 10.1186/1742-5573-6-4

Guloksuz, S., Pries, L. K., & van Os, J. (2017). Application of network methods for understanding mental disorders: Pitfalls and promise. *Psychological Medicine*, *47*(16), 2743–2752. doi: 10.1017/s0033291717001350

Hallquist, M. N., Wright, A. G. C., & Molenaar, P. C. M. (2021). Problems with centrality measures in psychopathology symptom networks: Why network psychometrics cannot escape psychometric theory. *Multivariate Behavioral Research*, *56*(2), 199–223. doi: 10.1080/00273171.2019.1640103

Hayes, A. M., & Strauss, J. L. (1998). Dynamic systems theory as a paradigm for the study of change in psychotherapy: An application to cognitive therapy for depression. *Journal of Consulting and Clinical Psychology*, *66*(6), 939.

Hernán, M., & Robins, J. (2020). *Causal inference: What if*. Boca Raton: Chapman & Hall/CRC.

Jones, P. J., Williams, D. R., & McNally, R. J.. (2020). Sampling variability is not nonreplication: A Bayesian reanalysis of Forbes, Wright, Markon, and Krueger. *Multivariate Behavioral Research*, *56(2)*, 249–255. 10.1080/00273171.2020.1797460.

Kendler, K. S., Aggen, S. H., Flint, J., Borsboom, D., & Fried, E. I. (2018). The centrality of DSM and non-DSM depressive symptoms in Han Chinese women with major depression. *Journal of Affective Disorders*, *227*, 739–744. doi: 10.1016/j.jad.2017.11.032

Kendler, K. S., Zachar, P., & Craver, C. (2011). What kinds of things are psychiatric disorders? *Psychological Medicine*, *41*(6), 1143–1150. doi: 10.1017/s0033291710001844

Kessler, R. C., Chiu, W. T., Demler, O., Merikangas, K. R., & Walters, E. E. (2005). Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the national comorbidity survey replication. *Archives of General Psychiatry*, *62*(6), 617–627. doi: 10.1001/archpsyc.62.6.617

Kim, P. J., & Jeong, H. (2007). Reliability of rank order in sampled networks. *The European Physical Journal B*, *55*(1), 109–114. doi: 10.1140/epjb/e2007-00033-7

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

Malgaroli, M., Calderon, A., & Bonanno, G. A. (2021). Networks of major depressive disorder: A systematic review. *Clinical Psychology Review*, *85*, 102000. doi: 10.1016/j.cpr.2021.102000

Martin, C., & Niemeyer, P. (2019). Influence of measurement errors on networks: Estimating the robustness of centrality measures. *Network Science*, *7*, 180–195. doi: 10.1017/nws.2019.12

Nesse, R. M., & Stein, D. J. (2012). Towards a genuinely medical model for psychiatric nosology. *BMC Medicine*, *10*(1), 5. doi: 10.1186/1741-7015-10-5

Pearl, J. (2000). *Causality: Models, reasoning, and inference*. New York, NY, USA: Cambridge University Press.

Petersen, M. L., Porter, K. E., Gruber, S., Wang, Y., & van der Laan, M. J. (2012). Diagnosing and responding to violations in the positivity assumption. *Statistical Methods in Medical Research*, *21*(1), 31–54. doi: 10.1177/0962280210386207

Piccinelli, M., & Wilkinson, G. (2000). Gender differences in depression. Critical review. *The British Journal of Psychiatry: The Journal of Mental Science*, *177*, 486–492. doi: 10.1192/bjp.177.6.486

Rehkopf, D. H., Glymour, M. M., & Osypuk, T. L. (2016). The consistency assumption for causal inference in social epidemiology: When a rose is not a rose. *Current Epidemiology Reports*, *3*(1), 63–71. doi: 10.1007/s40471-016-0069-5

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. doi: 10.1017/S0033291719003404

Schiepek, G., & Tschacher, W. (1992). Application of synergetics to clinical psychology. In W. Tschacher, G. Schiepek, & E. J. Brunner (Eds.), *Self-organization and clinical psychology: Empirical approaches to synergetics in psychology* (pp. 3–31). Berlin, Heidelberg: Springer Berlin Heidelberg.

Schwartz, S., Gatto, N. M., & Campbell, U. B. (2012). Extending the sufficient component cause model to describe the stable unit treatment value assumption (SUTVA). *Epidemiologic Perspectives & Innovations*, *9*(1), 3. doi: 10.1186/1742-5573-9-3

Schwartz, S., Gatto, N. M., & Campbell, U. B. (2016). Causal identification: A charge of epidemiology in danger of marginalization. *Annals of Epidemiology*, *26*(10), 669–673. doi: 10.1016/j.annepidem.2016.03.013

Steinley, D., Hoffman, M., Brusco, M. J., & Sher, K. J. (2017). A method for making inferences in network analysis: Comment on Forbes, Wright, Markon, and Krueger (2017). *Journal of Abnormal Psychology*, *126*(7), 1000–1010. doi: 10.1037/abn0000308

Teasdale, J. D. (1983). Negative thinking in depression: Cause, effect, or reciprocal relationship? *Advances in Behaviour Research and Therapy*, *5*(1), 3–25.

van Borkulo, C., Boschloo, L., Borsboom, D., Penninx, B. W. J. H., Waldorp, L. J., & Schoevers, R. A. (2015). Association of symptom network structure with the course of depression. *JAMA Psychiatry*, *72*(12), 1219–1226. doi: 10.1001/jamapsychiatry.2015.2079

VanderWeele, T. J. (2009). Concerning the consistency assumption in causal inference. *Epidemiology*, *20*(6), 880–883. 10.1097/EDE.0b013e3181bd5638.

VanderWeele, T. J. (2019). Principles of confounder selection. *European Journal of Epidemiology*, *34*(3), 211–219. doi: 10.1007/s10654-019-00494-6

VanderWeele, T. J., & An, W. (2013). Social networks and causal inference. In Stephen L Morgan (Ed.), *Handbook of causal analysis for social research* (pp. 353–374). Netherlands: Springer.

Wells, J. E., & Horwood, L. J. (2004). How accurate is recall of key symptoms of depression? A comparison of recall and longitudinal reports. *Psychological Medicine*, *34*(6), 1001–1011. doi: 10.1017/s0033291703001843

Wichers, M. (2014). The dynamic nature of depression: A new micro-level perspective of mental disorder that meets current challenges. *Psychological Medicine*, *44*(7), 1349–1360. doi: 10.1017/s0033291713001979

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*, 719490. doi: 10.3389/fpsyt.2021.719490

https://doi.org/10.1017/S0033291723000132 Published online by Cambridge University Press