To cite this article:

Blanchard, M. A., Contreras, A., Kalkan-Cengiz, R. B., & Heeren, A. (2023). Auditing the research practices and statistical analyses of the group-level temporal network approach to psychological constructs: A systematic scoping review. *Behavior Research Methods*, *55*(2), 767-787. https://doi.org/10.3758/s13428-022-01839-y

Auditing the research practices and statistical analyses of group-level temporal network approach to psychological constructs: A systematic scoping review

M. Annelise Blanchard [1,2], Alba Contreras[1], Rana Begum Kalkan[1,3], & Alexandre Heeren[1,2,4]

This is a preprint of a manuscript that is now accepted for publication in Behavior Research Methods (DOI: https://doi.org/10.3758/s13428-022-01839-y. Analysis code, data, and supplementary materials can be found in the accompanying Open Science Framework project: https://osf.io/8647n/

[1] Psychological Sciences Research Institute, Université catholique de Louvain, Louvain-la-Neuve, Belgium

[2] Belgian National Science Foundation (F.R.S.-FNRS), Brussels, Belgium

[3] Katholieke Universiteit Leuven, Leuven, Belgium

[4] Institute of Neuroscience, Université catholique de Louvain, Brussels, Belgium

Author Note

M. Annelise Blanchard https://orcid.org/0000-0002-9605-7022

Alba Contreras https://orcid.org/0000-0001-7292-8770

Rana Begum Kalkan https://orcid.org/0000-0002-3321-916X

Alexandre Heeren https://orcid.org/0000-0003-0553-6149

Correspondence concerning this article should be addressed to M. Annelise Blanchard. Psychological Sciences Research Institute, Université Catholique de Louvain, Place du Cardinal Mercier, 10, B-1348, Louvain-la-Neuve, Belgium. E-mail: marie.blanchard@uclouvain.be

To cite this article, please use the following reference:

Blanchard, M. A., Contreras, A., Kalkan, R. B., & Heeren, A. (2023). Auditing the research practices and statistical analyses of the group-level temporal network approach to psychological constructs: A systematic scoping review. *Behavior research methods*, *55*(2), 767-787.

Abstract

Network analyses have become increasingly common within the field of psychology, and temporal network analyses in particular are quickly gaining traction, with many of the initial articles earning substantial interest. However, substantial heterogeneity exists within the study designs and methodology, rendering it difficult to form a comprehensive view of its application in psychology research. Since the field is quickly growing and since there have been many study-to-study variations in terms of choices made by researchers when collecting, processing, and analyzing data, we saw the need to audit this field and formulate a comprehensive view of current temporal network analyses. To systematically chart researchers' practices when conducting temporal network analyses, we reviewed articles conducting temporal network analyses on psychological variables (published until March 2021) in the framework of a scoping review. We identified 43 articles and present the detailed results of how researchers are currently conducting temporal network analyses. A commonality across results concerns the wide variety of data collection and analytical practices, along with a lack of consistency between articles about what is reported. We use these results, along with relevant literature from the fields of ecological momentary assessment and network analysis, to formulate recommendations on what type of data is suited for temporal network analyses as well as optimal methods to preprocess and analyze data. As the field is new, we also discuss key future steps to help usher the field's progress forward and offer a reporting checklist to help researchers navigate conducting and reporting temporal network analyses.

Keywords: temporal network analysis, scoping review, ecological momentary assessment, time series data, network approach

Auditing the research practices and statistical analyses of group-level temporal network approach to psychological constructs: A systematic scoping review

The study of complexity has gained traction within multiple fields over the last few decades (Barabási, 2012), and psychology has recently embraced approaches exploring complexity as well. One such example is the network approach, which graphically models psychological phenomena as dynamic systems of interacting variables: a network (for an overview of the methodological approach, see Borsboom et al., 2021). Most of the research applying a network approach within psychology has specifically focused on psychopathology, viewing mental disorders as arising from the interactions between symptoms instead of from one common cause (Borsboom & Cramer, 2013). This view of psychopathology as a system of symptoms that interact and exacerbate one another is termed the network theory of psychopathology, and it comes with several crucial assumptions: that a central, or well-connected, symptom plays a crucial role in maintaining the entire psychopathology network, for example, or that a network with more connections between symptoms indicates more severe pathology (Borsboom, 2017). These assumptions are still under investigation (for reviews, see Blanchard & Heeren, 2022; Bringmann et al., 2022; McNally, 2021), but researchers have embraced the crucial tenet that psychopathology should be investigated by examining the complex interactions between symptoms—even outside the context of network theory. Indeed, the network approach within psychology has also included variables other than symptoms (Jones et al., 2017) and constructs other than psychopathology (e.g., personality: Cramer et al., 2012; intelligence: Van Der Maas et al., 2017; attitudes: Dalege et al., 2016). However, most psychology research involving network analyses thus far has utilized cross-sectional networks, and therefore mainly investigate the interactions between variables at one timepoint. Most psychological constructs, however, evolve over time, and examining relationships between

variables from one timepoint to the next is the next horizon in the network approach to psychology (Blanchard & Heeren, 2022; Bringmann et al., 2022).

Temporal network analyses are a specific type of network analysis estimated using intensive longitudinal data (e.g., data collected repeatedly over time), allowing researchers to visualize the temporal connections between psychological variables. These networks graphically model how variables interact and predict one another from one timepoint to the next (more so than just the past information from one variable alone, and so providing information about Granger causality; Jordan et al., 2020). Temporal networks thus have enormous potential to inform our theoretical insight into psychological constructs by investigating dynamic questions, such as how variables evolve or interact over time (Jordan et al., 2020; Wichers et al., 2017). For example, in a recent study on social media use and well-being, time spent on Facebook at one timepoint predicted negative affect, decreased self-esteem, feeling insecure and social comparisons at the next timepoint. In contrast, none predicted Facebook use, suggesting that the relationship between Facebook use and these variables is one-way and not bidirectional (Faelens et al., 2021). In another example examining eating disorder symptoms in at-risk populations, cognitions about fatness and fear of weight gain predicted many other symptoms, highlighting the central role of these variables in the development of eating disorders (Levinson et al., 2020). Both of these examples also illustrate the usefulness of temporal network analyses for clinical applications, by suggesting key variables or symptom inter-relations to target in interventions. Although most of the temporal network articles so far have focused on clinical populations or research questions, others have investigated questions involving general emotion dynamics (Bringmann et al., 2013; Elovainio et al., 2020; Martín-Brufau et al., 2020; Meng et al., 2020) or even personality (Lazarus et al., 2020; Pavani et al., 2017).

Temporal networks offer great potential to psychology, but they also require demanding data collection and novel analysis methods. Indeed, temporal networks are estimated on data with many timepoints such as from ecological momentary assessments (EMA)[1]: this data is typically collected in participants' daily lives, with participants answering short questions multiple times a day over several days or weeks (Myin-Germeys et al., 2018). Dynamic networks can also incorporate intensive longitudinal data that is not self-reported, such as from activity trackers (Nehrkorn-Bailey et al., 2018) or passive digital phenotyping (Lydon-Staley et al., 2019). The intensive longitudinal data is typically analyzed using multilevel vector autoregressive (mlVAR) models, which regress a variable at time *t* on that same variable at time *t-1* within a multilevel framework (to account for timepoints within subjects). However, other analyses to generate group-level temporal networks also exist. Examples include Bayesian multivariate multilevel models and group iterative multiple model estimation (GIMME), which builds a group network structure from subject-specific networks (for a clear overview of these different analysis methods, see Jordan et al., 2020; for an in-depth explanation, see Epskamp et al., 2018). For most temporal network models, one can estimate up to 3 networks, each visualizing a different type of relationship between nodes: contemporaneous (i.e., relations between nodes within one timepoint, controlling for temporal relationships; hypothesized to capture processes that occur faster than the timepoints are measured; Epskamp et al., 2018), temporal (i.e., relations between nodes from one timepoint to the next), and between-subjects (i.e., relations between means of different subjects). After generating networks, many researchers examine specific questions, such

[1] Although researchers use a variety of terms for intensive longitudinal data collection methods, including experience sampling methodology (ESM) and ambulatory assessment (AA), as well as daily diaries for repeated daily measures, we relied upon the term ecological momentary assessment (EMA) throughout the article, as this term is considered to have a broad methodology including many longitudinally intensive assessment traditions (Trull & Ebner-Priemer, 2009).

as which are the strongest edges (e.g., connections) between nodes, or which nodes are the most central (e.g., most strongly connected to other nodes, and therefore assumed to influence the entire network).

However, as the data collection and analysis methods are novel and constantly evolving, there is substantial heterogeneity in network study designs. This is also heightened by temporal network analyses being relatively new to psychology, with the methodology to compute temporal networks only recently introduced (Bringmann et al., 2013; Epskamp et al., 2018). Some of these varied practices also pertain to difficulties relevant to other fields (e.g., which assumptions do researchers check), while others are specific to temporal network analysis (e.g., varying methods to estimate and visualize temporal networks). Different research questions may also lead to different research practices: researchers could estimate just one network and examine specific edges or nodes, or they could compare networks between two groups (sometimes visually, sometimes with statistical analyses), sometimes focusing on differences in edges or nodes, sometimes in more global network topology such as connectivity. These varied approaches also lead to heterogeneous results that can be difficult to compare or interpret, and the field overall lacks a comprehensive view.

Network analyses have become increasingly common in psychology, and temporal network analyses, in particular, are quickly gaining traction, with many of the initial articles earning substantial interest from psychology researchers. Since the field is so quickly growing, we saw the need to formulate a comprehensive view of current temporal network analyses. More specifically, we wanted to form a clearer understanding of what choices researchers made when collecting, processing, and analyzing data, with the goal of helping future researchers make informed decisions when conducting their own temporal network analyses. To be clear, temporal networks are a new and complex methodology, and the purpose of this review is to identify areas

where the field as a whole can improve and to support coherence and consistency in how methods and analyses are reported. To systematically chart researchers' practices when conducting temporal network analyses, we reviewed articles conducting temporal network analyses on psychological variables (published until March 2021) in the framework of a scoping review. Our goal with this systematic audit of the field is to determine the common practices involved in conducting temporal network analyses, and thus identify trends and gaps (Paré et al., 2015). We also synthesize this data to provide guidelines and guidance to help future researchers using temporal network analyses in psychological science.

Methods

A scoping review maps the key concepts or practices in emerging fields, and identifies gaps in a research field (Peters et al., 2015). This is more suited to our goal of auditing the specific data collection and analysis practices for psychological studies conducting temporal network analyses than a systematic review, which strives to "answer a clinically meaningful question or provide evidence to inform practice" (Munn et al., 2018; for examples of scoping reviews auditing data collection and analysis practices, see Kjellberg et al., 2016; Zarin et al., 2017). We followed the PRISMA extension guidelines for scoping reviews (PRISMA-ScR; Tricco et al., 2018). Before beginning this scoping review, we preregistered our protocol, including our objectives, inclusion criteria, and methods on the Open Science Framework (OSF): https://osf.io/jwqmk/. Note that we slightly updated this registration (i.e., changes to the title and a few of the variables to be extracted). A complete list of the changes can be found here: https://osf.io/pgkr2/.

Search Strategy

We searched in two psychology databases: Scopus and PsycInfo. The initial search took place in January 2020, with two updated searches in March 2020 and March 2021. The exact

search strings used can be found in Figure 1. Although most of this research has been conducted within clinical psychology, we did not restrict the scoping review to a specific domain of psychological sciences to be as inclusive as possible and maximize generalizability from the audit. Figure 1 shows the exact search phrase for each database.

Inclusion Criteria

We limited our search to English-language studies that included data collected from human subjects and pertained to psychological variables. We limited our search to peer-reviewed articles to ensure that the methods and analyses reviewed in this scoping review were approved by experts. Eligible articles also had to include time-series data and investigate temporal network dynamics at a group level. Although there are also a number of articles examining idiographic (i.e., personalized) networks, we decided to limit our review to nomothetic (e.g., group-level networks), since the data collection, methods and analyses vary substantially.

Study Selection Process

Two reviewers independently inspected all titles and abstracts for eligibility according to the inclusion criteria. In a second phase, the criteria were reviewed in the full texts of the articles. In the case of inconsistencies, the two reviewers discussed with a third reviewer to resolve any remaining disagreements. We used CADIMA (Kohl et al., 2018) for the study selection process. Due to a technical issue with CADIMA, we only have an inter-rater agreement for the third wave of the study selection process (assessing 105 abstracts), with a kappa value of 0.61, which is considered “good.” This translates to disagreements on 18 abstracts, all of which were discussed and resolved with a third reviewer. Most inconsistencies were between “unclear” and “no” for criteria, reflecting that information about the inclusion criteria was not always apparent in the abstract.

Data Extraction (i.e. charting the data)

Data extraction included variables about metadata, sample characteristics, data collection, data preprocessing, estimation of data analysis, visualization of data analysis, robustness[2], and open science practices. We preregistered the exact variables for data extraction along with our protocol (https://osf.io/58swq/).[3] Three reviewers jointly developed the data extraction form to determine which exact variables to extract, and they piloted it before use on two randomly selected temporal network articles.

Two reviewers independently read all articles that met inclusion criteria and extracted the relevant information using a Google form (the contents of which can be found here: https://osf.io/mpj89/, and the link to the original form is in the preregistered protocol). The mean interrater agreement for categorical variables, calculated using the same procedure as (LópezNicolás et al., 2021) was a Cronbach's kappa of k = .67 (SD = .25), signifying substantial agreement. The range of agreement across variables was quite wide, however, with kappa's ranging from .08 to 1, reflecting how difficult it was to determine specific information for many of the variables investigated, in many cases because the articles themselves did not clearly describe all details of the data collection or analysis process. Only a few variables had a kappa below .5 (see Table S1), and we present more information regarding the Cronbach kappa's for the categorical variables in the supplementary materials. All disagreements between the two reviewers were extensively reviewed and discussed with a third reviewer.

Synthesis of Results

[2] By robustness, we refer to "the stability of statistical inference under the variations of the accepted distribution models" (Shevlyakov & Vilchevski, 2002, p. *vii*); in other words, whether the model estimation yields comparable results with slight variations of the data (e.g., bootstrapped samples, case-dropping).

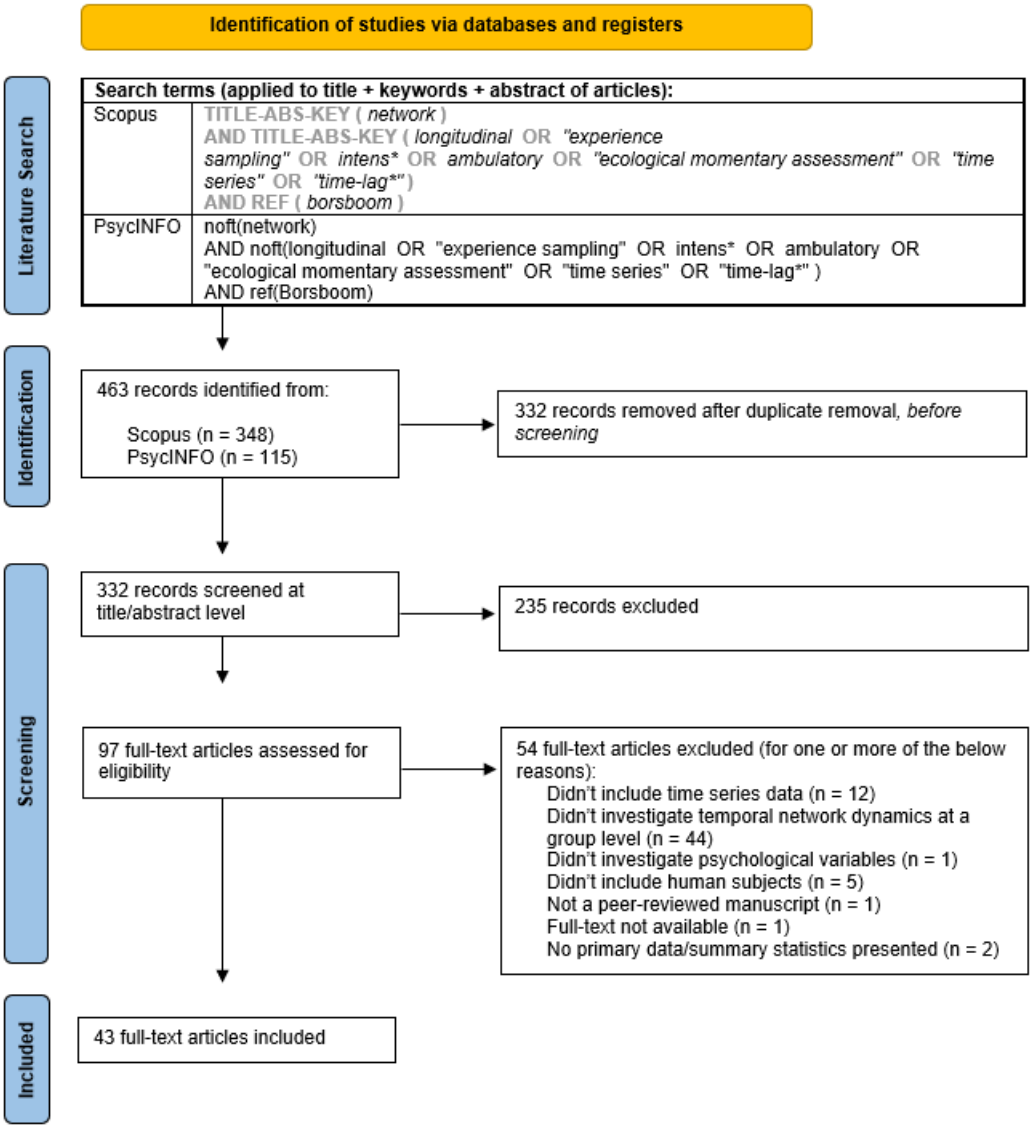
[3] The original preregistered list of variables for data extraction can be found here: https://osf.io/v73nb/. We updated our preregistration in March 2021 when adding a year onto our search end date, and we also added five additional questions to our data extraction form, which we found pertinent after reviewing the first batch of articles. These changes are detailed in blue in our updated preregistration (https://osf.io/pgkr2/), and have to do with authors computing multiple networks in their paper, the specific model they use, and how the authors handle missing data.

The full dataset is available on OSF as a .csv file (https://osf.io/jsqrc/). For convenience, we also share a web-based interface that can be used to filter or group any variable of the full dataset in an easier fashion (https://airtable.com/shrcC5MOBhHwEbwD8). In this review, we also summarize key information relating to the data collection, preprocessing, data analysis, open science, and robustness practices. We present graphs for key results that were created using R, adapted from the R code of López-Nicolás et al. (2021).

Results

The PRISMA flow diagram in Figure 1 illustrates the study selection process for this scoping review. For each extracted variable, readers can see the full list of possible answers for categorical variables (e.g., answered via multiple choice option by the two extractors) and the exact wording for free-response variable in the data extraction form: https://osf.io/mpj89/..

Figure 1. Literature Search Process



Note: We included a search term for their reference section of all articles: articles had to reference D. Borsboom, since he has authored the main initial conceptual papers linking network analysis with psychological research; we assume that any article discussing the psychological implications of temporal network analyses would cite at least one of Borsboom's initial articles. We included this search term to exclude articles with similar keywords that were not relevant to our search, such as research involving brain imaging, social networks, ecology, and so on. Figure template from (Page et al., 2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. BMJ, 372, n71. https://doi.org/10.1136/bmj.n71. For more information, visit: http://www.prisma-statement.org/.

Study & Sample Characteristics

Eighty-eight percent of articles (*n* = 38) were either part of a larger project or a reanalysis of already preexisting data. The sample size per network ranged immensely (see Table 1).

Regarding the type of sample, 51% of articles had a clinical or health-related sample (n = 22), 33% of articles had a non-clinical (i.e., healthy) sample (n = 14), and 16% of articles included both clinical and non-clinical participants (n = 7).

Table 1
Descriptive Information about Network Variables

Variable	*M*	Median	*SD*	Min	Max
Number of participants	119.6	66	210.7	5	1354
Total possible number of timepoints	60	61.2	21.3	7	140
Mean number of completed timepoints	47.9	47.5	22.8	13.8	114.8
Number of nodes	9	7	4.9	4	30

Data Collection

Timescale

A crucial point for any study assessing variables over time is the timescale, and the articles varied enormously in the frequency of assessment, from ten times a day to once a year. However, the vast majority of articles (81%, n = 35) used a daily or multiple-times-a-day frequency of assessment, over a period of time ranging from five days to 56 days. We report the exact assessment schedule of each study in Table S2 in the supplementary materials.

Number of Timepoints & Missing Timepoints

Another key aspect of data collection is the number of timepoints, which also varied widely (see Table 1). Sixty-seven percent of articles (n = 29) also reported the mean number of completed timepoints (i.e., not including the timepoints that participants missed answering or that were excluded during analyses; see Table 1). Finally, the scoping review also revealed that temporal network research takes a variety of actions in response to participants missing timepoints: 39.5% of articles (n = 17) do not exclude participants based on missing timepoints, 18.6% of articles (n = 8) do not specify any action, while 41.9% of articles (n = 18) do exclude participants if they miss answering a certain percentage of timepoints. More specifically, 23.3% of articles (n = 10) excluded participants if they did not complete at least 30% of timepoints,

4.7% (n = 2) excluded participants not completing at least 40% of timepoints, 9.3% of articles (n = 4) excluded participants without at least 50% of valid responses while 4.6% of articles (n = 2) exclude participants not completing more than 50% of total timepoints.

Data Collection Method & Design

We also examined details regarding how the researchers collected data in the reviewed articles: 77% of articles (n = 33) relied on an electronic device, 9% (n = 4) used a combination of electronic device and pen and pencil, and 14% (n = 6) did not specify. The most frequent data collection designs (categorized following the descriptions provided by Myin-Germeys et al., 2018) were a fixed sampling design, with questionnaires completed at equal time intervals (33% of articles, n = 14) and a pseudo-random sampling schedule, with questionnaires completed at random times within a specific time interval (44% of articles, n = 19). Furthermore, to increase participants' compliance with the questionnaires, 49% of articles (n = 21) compensated participants for their time. More information about the exact devices and platforms, as well as data collection designs and compensation details, can be found in the Data Collection section of the supplementary materials or in the full dataset.

Data Preprocessing

In 98% of studies, authors used R to conduct analyses[4]. Most studies reported that the variables they included in their networks were ordinal (Likert scale) variables (79% of studies) and that nodes were composed by using only one item (93% of studies). The number of nodes varied widely (see Table 1). For more information about the statistical software and packages the

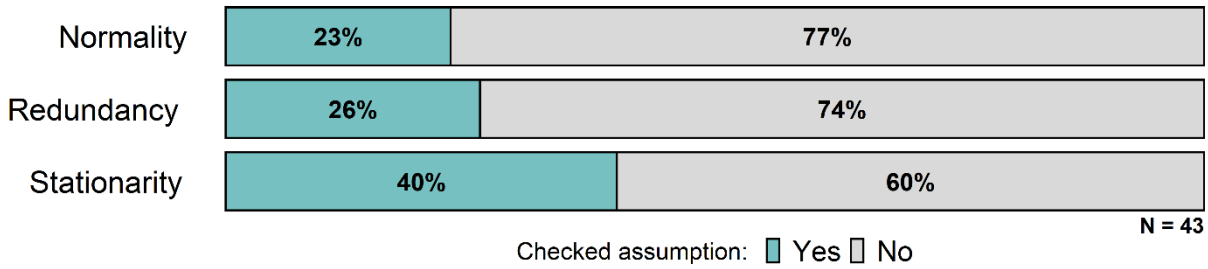
[4] Several authors reported using a combination of software types, but all but one used R for at least one portion of their analyses. For many other variables as well, authors sometimes utilized multiple of the answer options so that the total would amount to over 100%.

authors report using, as well as the type of variables assessed and the methods of defining nodes, see the Data Preprocessing section of the supplementary materials.

Assumptions

A key point before starting analyses is examining if there are violations of relevant assumptions, but only a minority of articles assessed whether their data was normal and stationarity (see Fig. 2). Table 2 lists the main methods used to assess and correct normality and stationarity, while more detailed explanations and references to specific articles can be found in the Assumptions Section of the supplementary materials.

Figure 2. Assumptions



Another step authors could verify before beginning analyses is whether nodes overlapped in content. Only 11 articles report checking for potential redundancy in node content (see Fig. 2); the methods used are listed in Table 2, while more information can be found in the Node Redundancy section of the supplementary materials.

Table 2
Specific Methods Used to Assess Data or Network During Temporal Network Analyses

Action	Specific Methods
Checking normality (n=9)	• Visually inspecting histograms (n = 3) • Kolmogorov-Smirnov test (n = 3) • Shapiro-Wilk test (n = 2)
Correcting for normality (n=3)	• Log transformation (n=2) • Transformation via the normal quantile distribution (n=1)
Checking stationarity (n=17)	• Kwiatkowski-Phillips-Schmidt-Shin (KPSS) unit root test (n=11) • Dickey-Fuller test (n = 1) • Visual inspection of participants' responses over time (n = 2) • Add time variable (and compare model fit; n=2; or see whether variables depend on this time variable; n=2) • Theoretically assume violation (e.g., participants in treatment, n=1)
Correcting for stationarity (whether or not first checked; n=18)	• Detrending the data (n = 10) • Including a linear trend (n = 8)
Checking redundancy (n=12)	• Choose nodes on a theoretical basis so they don't overlap (n=1) • Use model robust against high overlap in variables (e.g., PCMCI; n=1) • Look for multicollinearity or high correlations between variables (n=6) • Factor analysis (n=2) • Use the goldbricker function to identify which nodes likely measure the same underlying construct (n = 2)
Missing data (n=8)	• Taking specific action (n = 5): Multiple imputation; imputation; estimating data with Kalman filter; ensuring equal time intervals between timepoints (using cubic spline interpolation) • Justify that model is robust against missing data (n=3): because of Bayesian estimation, full-information maximum likelihood, or model itself
Network comparison (using formal statistical methods; n = 15)	• Using permutation tests (n = 10) • Using t-tests (n=3) • Using a Bayesian network model with 95% credible interval (n=1) • Correlating networks indices (to compare similarity of samples; n=1)
Robustness (n=6)	• Bootstrap confidence intervals (n=3; difficult with multilevel models) • Credible interval (uncertainty estimate with Bayesian estimation; n=1) • Case-dropping methods to assess stability (n=2)
Sensitivity (n=14)	Focus: • Variable transformation (e.g., standardization, to correct stationarity or normality violations; n=8) • Model choice or model specifics (n=4) • Node construction or selection (n=2)

- Inclusion of specific participants (n=4)

Note. More details (including the full list of articles) for each method can be found in the supplementary materials or in the full dataset (see https://airtable.com/shrcC5MOBhHwEbwD8). The number of articles utilizing a specific method does not always sum up to the number of articles taking a specific action, since 1) not all authors specified how they conducted a specific action (e.g., some authors stated their data was not normal but did not explain how they assessed this) and 2) some authors did multiple types of actions (e.g., many authors did multiple sensitivity analyses). PCMCI = Peter and Clark Momentary Conditional Independence.

Missing Data

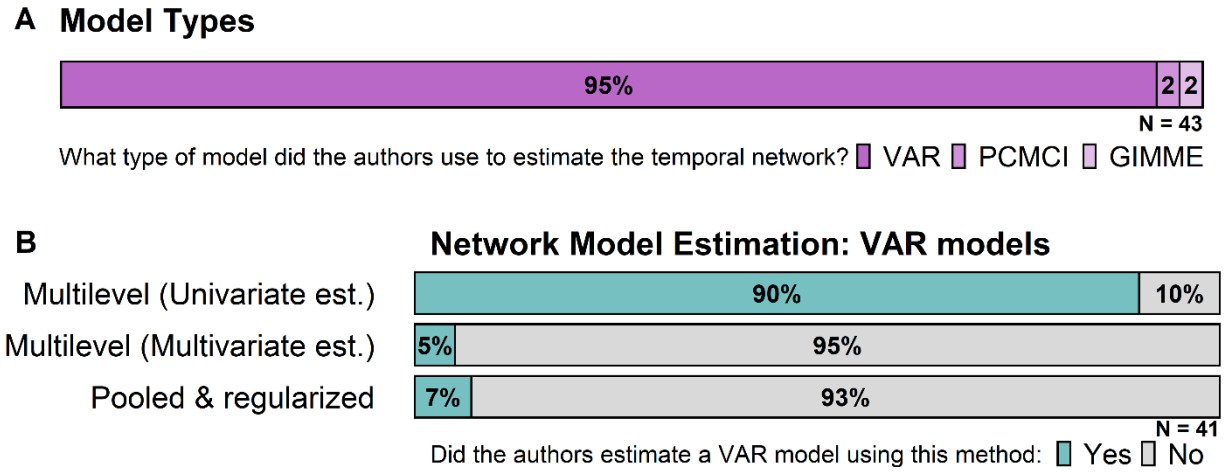
Regarding missing data, only five articles (12%) report any specific actions they took regarding missing data; the specific methods are listed in Table 2 (with more information and references in the supplementary materials). Since most of the models rely on mlVAR, most models therefore use the automatic action within the mlVAR package to use listwise deletion to deal with missing data (Epskamp et al., 2019); two articles explicitly mentioned this.

Data analysis: Estimation

Models

Within all articles, researchers estimated a group-level temporal network; 51% of articles (n = 22) also included a group-level contemporaneous network, and 36% of articles (n = 17) also included a between-subjects network. Almost all articles reported using a vector autoregressive (VAR) model to estimate the temporal network; the only other types of models reported were a GIMME model and a Peter and Clark Momentary Conditional Independence (PCMCI) model (see Figure 3A). The vast majority of VAR models were multilevel and estimated through sequential univariate estimations, although 2 articles used multivariate estimation (see Figure 2B). Only 2 articles did not use multilevel models (Lazarus et al., 2020; McCuish et al., 2021), but instead pooled data and then regularized it, with one additional article (de Vos et al., 2017) comparing a multilevel VAR model to a sparse VAR model (see Figure 3B).

Figure 3. Models



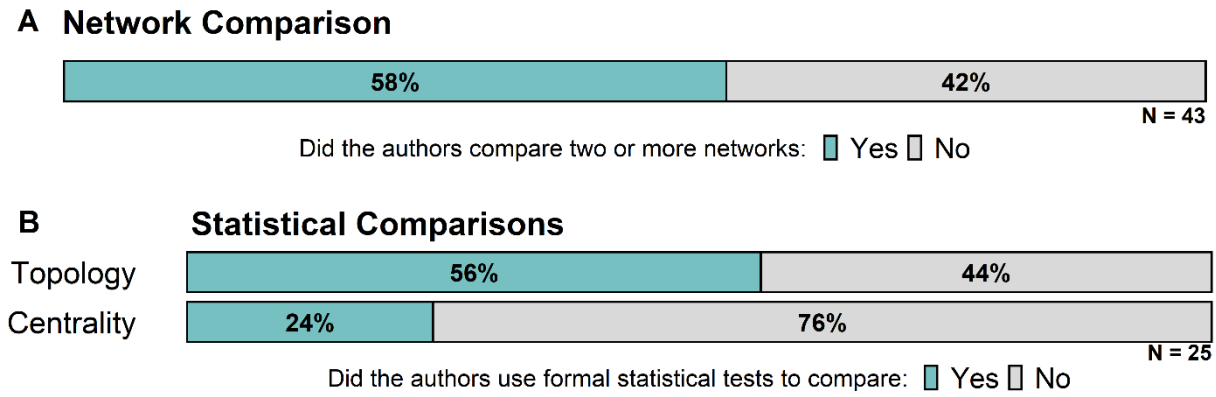
Note. All percentages reported in the figure are rounded. VAR = Vector autoregressive; PCMCI = Peter and Clark Momentary Conditional Independence; GIMME = group iterative multiple model estimation; est. = estimation.

All models used a lag of 1 (i.e., from time t-1 to time t), although three articles did run alternate models with other lags (t-2 or t-3) and chose the model with the best fit, which was always t-1 (Bringmann et al., 2013, 2015; Kaiser & Laireiter, 2019).

Network Comparison & Centrality

More than half of the articles estimated at least two networks and compared them in some manner, including visually (see Figure 4A); around half of these compared the topology of networks, and almost a fourth compared centrality indices (see Figure 4B). Table 2 lists the specific methods used, and more information can be found in the Data Analysis Section of the supplementary materials.

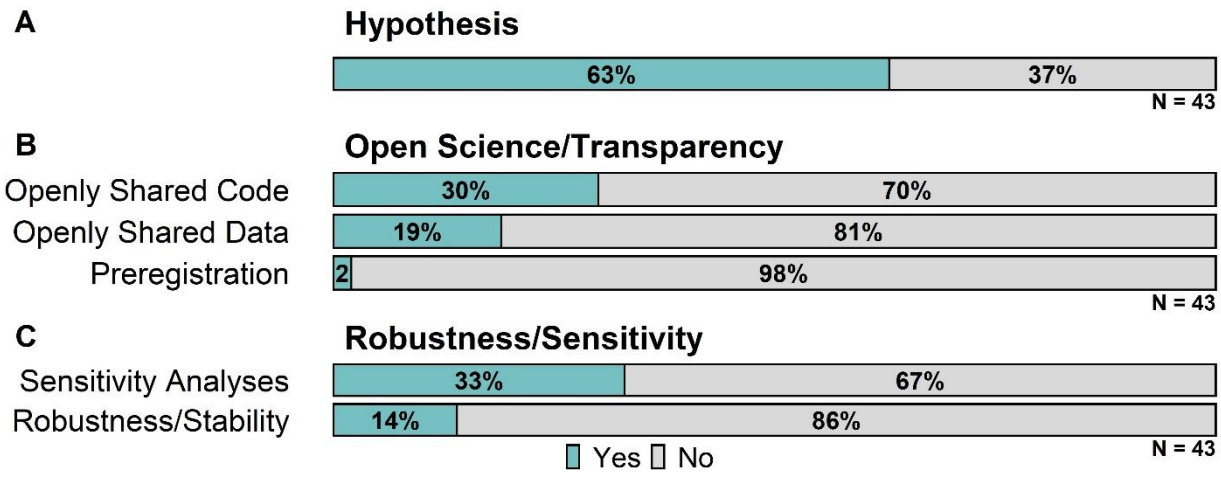
Figure 4. Network Comparisons



Open Science Practices

Only one article (Groen et al., 2020) preregistered their hypotheses and/or analyses for their network analyses, despite 61% of articles including specific hypotheses (see Fig. 5A). Few articles openly (e.g., include a link within the article itself) share their code and fewer the data (see Figure 5B). None of the studies were replications of previous studies (i.e., investigating the same construct), but two studies included validation samples within the same article to replicate their own results (Bringmann et al., 2013; Huckins et al., 2020).

Figure 5. Open Science and Robustness Practices



Robustness and Sensitivity

Only a few articles assessed the robustness of their edge or centrality estimates (see Fig 5C), using either Bayesian estimation or case-dropping bootstrap methods to do so (see Table 2, with more details in Table S3 of the supplementary materials).

One-third of the articles included sensitivity analyses around a specific analytical choice to observe its impact on the results (i.e., running the analysis with and without that choice; see Fig 3B). These articles performed sensitivity analyses on the impact of variable transformation, model or node specifics, and so on (see Table 2).

Discussion

Key Findings

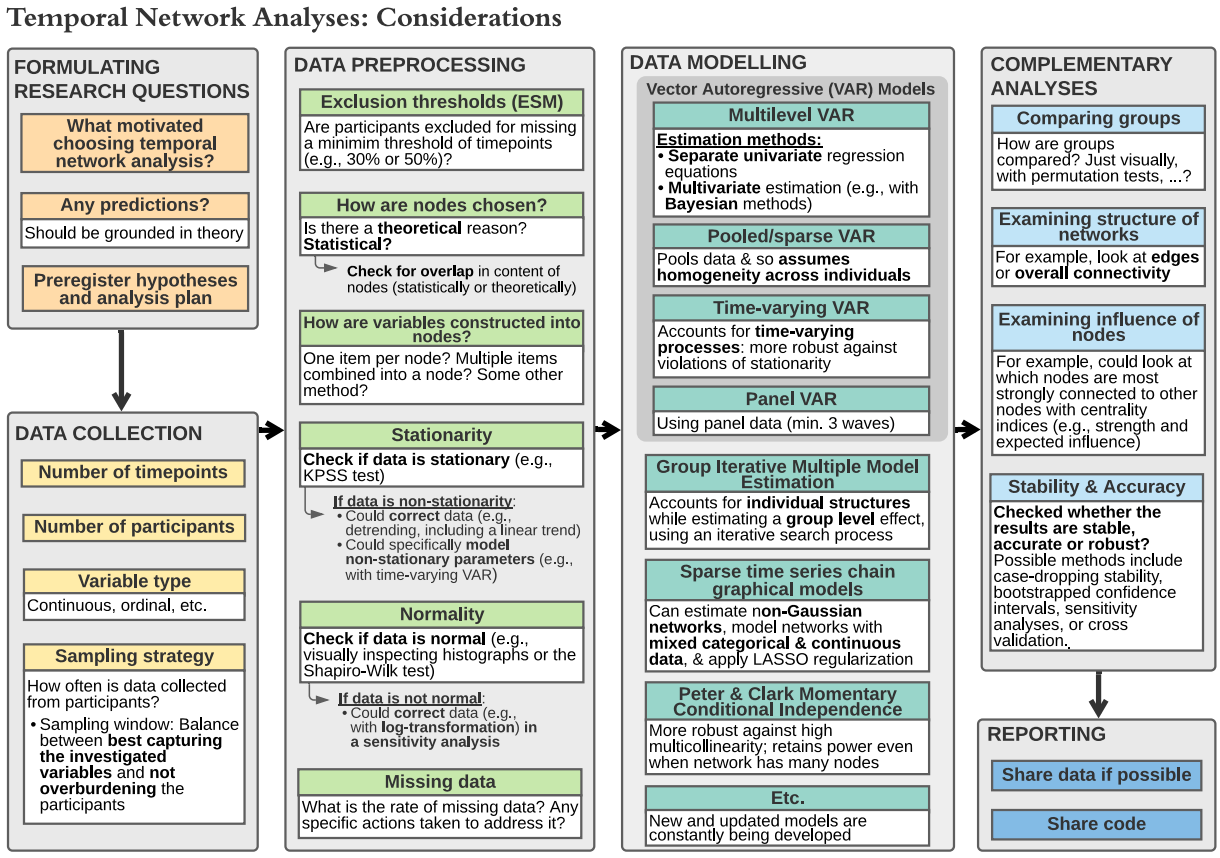
A striking observation throughout our audit of the methodology and analysis of temporal networks is that there is substantial heterogeneity in almost all aspects, from whether and how articles report node development and selection and preprocessing choices to how articles deal with missing data and assumption violations. This heterogeneity is reflected in the relatively modest interrater reliability ratings, both for article selection and for data extraction: there is no standard way of reporting information (about data collection or analyses), nor standard set of information to report. This is not unusual for a novel and growing field. It echoes our original reason to conduct a scoping review on this field: to understand what are the current main practices in data collection and analyses for temporal network analyses, so as to help interested researchers figure out what options they have when performing their own temporal network analyses.

Recommendations

We thus present Figure 6, which illustrates the main considerations involved when planning and conducting a temporal network analysis. The recommendations discussed are guided both by the literature, with relevant references cited, as well as by the results of this

scoping review. We also discuss these recommendations in greater depth in the following sections. We hope readers interested in conducting temporal network analyses can use this information to help them make and justify decisions about collecting, preprocessing, and analyzing data. We also provide these considerations in the form of a reporting checklist, to help authors ensure they are reporting all relevant details in their manuscript and/or supplementary materials; this form can also be used as a preregistration checklist (https://osf.io/e6wp3/).

Figure 6. Temporal Network Analyses: Considerations



Note. The considerations in this figure were formulated using the extracted data from this scoping review as well as relevant literature (citations in the discussion section); more detailed explanations for all aspects of this figure can be found in the discussion section. This information is also presented as a reporting or preregistration checklist here: https://osf.io/e6wp3/. KPSS = Kwiatkowski-Phillips-Schmidt-Shin (KPSS) unit root test; LASSO = least absolute shrinkage and selection operator.

Formulating Research Questions

Before starting a temporal network analysis, researchers should ensure that they have thoroughly developed their research question(s), and that these research questions are suited to temporal network analysis. Temporal network analyses are ideal for investigating how dynamic variables interact with one another over time; researchers can also more specifically investigate which edges or nodes are especially important, or how the temporal networks of two different groups compare. It is crucial, though, that researchers ground the justifications and predictions of their network analysis in theory. At the most basic level, this means that researchers should use temporal network analysis techniques because investigating temporal dynamics in a multivariate fashion will shed light on heretofore unanswered questions in their topic of interest. This also means researchers should use theory to ground their choice of nodes. For example, Faelens et al. (2021) sought to understand the relationship between using social network sites (SNS) and well-being. Previous cross-sectional research had identified several indicators that played a role in the relationship between well-being and SNS use, including social comparison, self-esteem, and repetitive negative thinking. Moving to a temporal network perspective allowed these researchers to investigate the direction of relationships between all of these variables. Theory should also justify hypotheses. For example, Groen et al. (2020) investigated the temporal dynamics between depressive and anxiety disorders, and made hypotheses based on network theory that overlapping mental states would function as bridges connecting the two disorders (e.g., activating symptoms in the other disorder, leading to comorbidity).

Despite the great efforts of research to include theoretically relevant variables in temporal models, the vast majority so temporal network analyses thus far are data-driven: authors investigate a potential network structure by matching a statistical (network) model to data. Although some of the reviewed articles include specific hypotheses, none are derived from a

testable and falsifiable theoretical network model. This would involve developing a theoretical network model of the psychological construct of interest: identifying the nodes in the network, as well the dynamic associations between nodes (e.g., how changes in one node might propagate throughout the network), and thereby developing testable (and importantly, falsifiable) hypotheses (Bringmann et al., 2022). Bringmann and colleagues (2022) note that there are few of these theoretical network models; examples include a network theory of how attitudes change (Dalege et al., 2016) and a computational model of panic disorder (Robinaugh et al., 2019). Testable hypotheses derived from theoretical network models can be examined through fitting statistical network models to data (potentially through "confirmatory" network models which test model structure and fit; Epskamp, Rhemtulla, et al., 2017), although also through other designs (e.g., experimental). Network analysis is still a new field, and using data-driven methods to explore plausible network structures is also a worthwhile pursuit, but hopefully there will be more theory-driven and confirmatory approaches to complement the current exploratory nature of temporal network analyses.

Preregistration

We urge researchers to preregister their planned preprocessing and analytical steps, as well as their data collection plans (or knowledge of preexisting data if conducting a reanalysis; Weston et al., 2019). This is especially important since so many of the preprocessing and analytical choices authors can make are arbitrary (e.g., excluding participants for missing a specific threshold of timepoints, constructing nodes, correcting assumption violations), and de Vos et al. (2017) firmly demonstrate that arbitrary analysis choices (e.g., definitions of density/connectivity, data transformation) can influence results.

We also encourage researchers to preregister any hypotheses they develop. Over half of reviewed articles did include predictions of some type, but only one (Groen et al., 2020)

preregistered its hypotheses and analysis plan. We note that many studies are exploratory and designed to generate hypotheses for future research; in this case, the authors should clearly state this in the manuscript itself (and in the potential preregistration). If researchers do have hypotheses, though, we would recommend that they preregister them.

Data Collection

Number of Timepoints. When collecting data, it is necessary to have enough timepoints per person. The median number of timepoints in the reviewed articles was 60. Within both the EMA literature (Palmier-Claus et al., 2011) and for most statistical models (Jordan et al., 2020), the absolute minimum number of timepoints suggested is 20, although this is typically the minimum number of timepoints the model needs to run (e.g., Epskamp et al., 2019). We recommend including many more timepoints, perhaps around 60 (the mean in this scoping review and another large EMA meta-analyses; e.g., Vachon et al., 2019). In general, the more timepoints, the better.

However, alternatives exist if a dataset has very few timepoints (e.g., less than 20 timepoints per person), such as using network models made for panel data (i.e., with at least 3-4 waves; Epskamp, 2020). Other alternatives include pooling data (such as by using pooled VAR to generate networks; Lazarus et al., 2020; McCuish et al., 2021); however, de Vos et al. (2017) caution that this method should only be used if there is little heterogeneity among participants' responses, since it assumes each participant is a replication of all others.

Number of Participants. The median number of participants in reviewed articles is 66 per network. In general, including more participants indicates greater power, but actual power analyses are complicated with temporal network models. It is possible to conduct simulations and vary the number of subjects, nodes, and timepoints (this is integrated within the mlVAR package, for example), but the researcher should know *a priori* the network structure (Jordan et al., 2020).

However, if researchers would like to explore simulating different numbers of participants and timepoints while taking into account temporal dependence (for analyses such as multilevel regression models, if not explicitly for temporal network analyses), Lafit et al. (2021) have recently developed a Shiny app and tutorial for doing so.

In general, though, researchers should know that power for longitudinal data used in multilevel models takes into account both number of timepoints and number of participants (Jordan et al., 2020) and is also dependent on the effect size, type of model, and so on (Bringmann et al., 2022). In addition, number of nodes is relevant to this calculation: a network with more nodes includes more parameters to be estimated, thus requiring more timepoints and ideally more participants.

Variable Type. Continuous variables allow more variability in participant responses. Aalbers and colleagues (2019), for example, use a scale of 0-100 to prevent a restricted range issue (i.e., a variable with minimal variance), which can distort conclusions made from network models (e.g., Terluin et al., 2016). We therefore recommend a scale with a large range, such as from 0-100, since other scale types (e.g., Likert) could yield limited variability in participant responses since they include fewer response options.

The most common variable type, used in 79% of reviewed articles, were ordinal variables (specifically, Likert scale). From a statistical standpoint, it is optimal if all variables have the same scale. Binary variables can be challenging to implement with current network models (although Klippel et al., 2018, conducted sensitivity analyses using multilevel logistic regressions).

Sampling Strategy. Reviewed articles included sampling strategies ranging from 10 times per day over a few days to once a month for four years. In general, the sampling window should be selected to best capture the investigated variables while also not overburdening

participants or leading to missing timepoints; finding the correct balance often takes piloting (Varese et al., 2019). The frequency of sampling (e.g., many times a day, twice a day) should be chosen to best capture the temporal variations of the variable of interest (Trull & Ebner-Priemer, 2020). For example, if a variable fluctuates intensely throughout the day, such as with emotional states, it would likely be better to include many measurements throughout the day. If, on the contrary, a variable fluctuates at a slower rate, such as sleep quality, it might make more sense to measure daily (Dejonckheere & Erbas, 2021). In addition, researchers should keep in mind that they might need to specify in their R code if they are measuring participants multiple times a day for some packages (e.g., mlVAR).

It is typically simpler for most of the models used in the reviewed articles (e.g., models such as multilevel VAR, that model time in a discrete fashion instead of continuously) to have all variables assessed on the same timescale (e.g., all assessed daily). However, this might not be possible if the variables of interest occur at different timescales (e.g., if investigating both sleep and emotions) or if the variables only occur occasionally (e.g., panic attacks); in these cases, continuous time dynamic models could be employed, as they grant greater flexibility with timescales (for an overview, see Bringmann et al., 2022).

Data Preprocessing

Exclusion Thresholds Based on Participant Response Rate. The EMA literature recommends that each participant have responses for at least 20 timepoints (Palmier-Claus et al., 2011). Twenty-one percent of reviewed articles excluded participants if they did not answer a minimum threshold of EMA items, most commonly 30% of items or less than 20 timepoints.

We recommend a flexible approach. For instance, if researchers collect 200 timepoints, a participant who only answered 30% of prompts still answered 60 timepoints. It could therefore make more sense to establish a minimum number of timepoints, such as only including in

analyses participants who answered at least 20 timepoints (often considered the absolute minimum number of timepoints needed for analyses such as VAR models to run) or some other number of timepoints. It is also worth noting that participants with higher amounts of missing data, or even attrition, in longitudinal research may be those with higher symptom severity (e.g., Abdullah et al., 2021; Lamers et al., 2012), and so it can bias the results to exclude them.[5]

Node Selection. Researchers should carefully consider and justify their choice of nodes – are they chosen for theoretical reasons? Statistical ones? Typically, researchers should choose which variables they include as nodes based on theoretical background, as discussed in the 'Formulating research questions' section with the example of Faelens et al. (2021). However, researchers might decide not to include specific nodes for statistical reasons, such as if they show no temporal fluctuations. If possible (e.g., if the researchers are collecting the data themselves), all variables pertinent to the research question should be included in the network without extraneous variables (Bringmann et al., 2022), and these variables should be assessed in a manner suitable to frequent assessment.

In general, an increase in the number of nodes in a network goes along with an increase in model complexity. In reviewed articles, the number of nodes ranges from four to 30. Researchers should check for overlap in the content of nodes, either statistically or theoretically. If high overlap is unavoidable for theoretical reasons (e.g., studying a construct which has highly overlapping features), researchers could use a model that allows many nodes and highly correlated variables (such as the PCMCI model; Runge et al., 2019) or a model that explicitly takes into account the overlap between nodes (e.g., a latent network model; Epskamp, Rhemtulla, et al., 2017).

[5] We are thankful to one of the reviewers of this manuscript for pointing our attention to this issue.

Variable Construction into Nodes. Researchers should make sure to report how they constructed nodes from their variables. For example, did each item form a node? Were multiple items combined into a node, and how? Was some other method used?

Stationarity. Since stationarity is one of the main assumptions of most models with EMA data, researchers should check if their data is stationary (e.g., if the mean and variance of the EMA data remain unchanged; Jordan et al., 2020). One way to do so is to use the KPSS unit root test (Kwiatkowski et al., 1992), as 11 reviewed articles did. If data is non-stationary, researchers could correct this by detrending the data (e.g., Hoffart & Johnson, 2020) or including a linear trend (e.g., Bringmann et al., 2015). However, one may wonder about the influence of these types of corrections on the interpretability of the results (Jordan et al., 2020). Therefore, researchers could instead directly use a model that allows non-stationary data by specifically modeling non-linear time-varying parameters, such as time-varying VAR (Haslbeck et al., 2020). Perhaps researchers are investigating processes that change over time (e.g., during treatment, or a transition into or recovery from a mental disorder); in this case, researchers could use models that explicitly look for abrupt or gradual changes, although these models typically require many more timepoints (for an overview, see Bringmann et al., 2022).

Normality. Another crucial assumption for most temporal network models (e.g., VAR models; Epskamp, Waldorp, et al., 2018) is (multivariate) normality, and so researchers should check whether the assumption of normality is violated. They can do so by visually inspecting histograms of the residuals or with the Shapiro-Wilk test (Shapiro & Wilk, 1965).[6]

If data is non-normal, it is currently unknown exactly how this will bias temporal network results if using an estimation technique that assumes normality (Epskamp, 2020), especially since

[6] Although the Komogorov-Smirnov test was also used by three reviewed articles, the Shapiro-Wilkes test is favored as less biased and more powerful (Razali & Wah, 2011; Steinskog et al., 2007)

violations of normality can occur for many different reasons (Epskamp et al., 2018). For example, if data is non-normal because it was measured on a different scale (e.g., Likert-type ordinal scale) but the underlying process is normal, it can be effective to transform the data back to normal (Epskamp et al., 2018). For instance, among reviewed articles, this was carried out by de Vos et al. (2017), who transform all items using the normal quantile transformation, since these items had Likert-scale responses with skewed distributions; they also report differences in network density compared with non-transformed data (among other sensitivity analyses). However, data that is non-normal due to the process itself being non-normal (e.g., if the residuals are skewed) pose more of a problem (Epskamp et al., 2018).

One possible action when faced with non-normal data is to transform the data, although how transforming time-series data affects temporal networks is, to the best of our knowledge, currently unknown. Among the reviewed articles, in addition to de Vos et al. (2017), only two additional articles performed a transformation for non-normal data. First, Aalbers et al. (2019) report the distributions of their variables, as well as the normality of within-person means; upon the request of a reviewer, they then log transformed items that had a skew or kurtosis outside the acceptable range of -2 to 2, if that transformed reduced the skew and kurtosis to within that acceptable range. They state that the overall pattern of results was unaltered by transforming the data (and only report the networks generated from the original non-corrected data). Second, Faelens and colleagues (2021) report that some variables had a non-normal distribution (as did their within-person means). They therefore include in their supplementary analyses a sensitivity analysis where they generate networks from log-transformed variables that had unacceptable ranges of skew and kurtosis. Faelens et al. (2021) state that the transformed and non-transformed networks show similar relationships between variables; and upon visually comparing the networks, the edges that are additional or missing are typically the thinnest edges. Possibilities

for transforming non-normal data other than log-transforming exist, such as the non-paranormal transformation sometimes used in cross-sectional network studies to yield multivariate normal data (Epskamp et al., 2018), which has been shown through simulation studies to be helpful when generating cross-sectional networks from continuous skewed data (Isvoranu & Epskamp, 2021). However, it is unknown how this transformation, applied to all variables at once, might affect the temporal dependencies present in time series data, or how it might impact the resulting temporal network estimation.

Since there is so little currently known about the impacts of violations of normality when generating temporal networks or transforming time-series data, the most we can recommend is that researchers check and clearly state whether the assumption of normality is upheld or violated. If it is violated and authors choose to transform their data in some manner, we recommend they do so as a sensitivity analysis and compare results with and without data transformation (which could briefly be described in the main manuscript but more fully included in the supplementary materials, for example), at least until more thorough investigations into non-normality, data transformation, and temporal networks are carried out. We also recommend that especially since the impact of normality violations are not yet fully understood, researchers could also check the robustness of their network results (e.g., by using case-dropping bootstrap techniques; Epskamp, 2020).

Unequal Time Observations. It is recommended to insert missing values when there are unequal time intervals between observations (or potentially impute missing data) to create a dataset with (roughly) equidistant timepoints (Hamaker et al., 2018). This is because some models (e.g., VAR models) assume equidistant data, or roughly equal time intervals between any two consecutive assessments (Jordan et al., 2020). One of the reviewed articles (i.e., Kaiser & Laireiter, 2019) responded to this model assumption by using cubic spline interpolation to

generate equally spaced time-series. Another possibility, especially if investigating variables that operate on different timescales (e.g., behaviors and affect) is to use models that do not assume equal time intervals between timepoints, such as continuous time dynamic models (for more information, see Bringmann et al., 2022).

Missing Data. Models estimated with mlVAR automatically use listwise deletion with missing data (Jordan et al., 2020), but other possibilities include imputing missing data (e.g., McCuish et al., 2021) or estimating missing data using a Kalman filter (e.g., Levinson et al., 2018). Two articles specifically used multiple imputation, which involves generating multiple imputed datasets, performing the required analyses with them, and then averaging the results: de Vos et al. (2017) specify they impute 10 datasets and averaged the subsequent results, and Levinson et al. (2020) use multiple imputation but do not further specify their methods. Some estimation methods are also more robust against missing data, such as full-information maximum likelihood (only when data is missing at random, however; Cham et al., 2017) or Bayesian estimation (e.g., Groen et al., 2020).

Ji et al. (2018) recommend against listwise deletion for missing data in intensive longitudinal assessment, as it is likely to lead to biased results, and instead suggest using multiple imputation (including partial multiple imputation supported by full-information maximum likelihood) and averaging results using Rubin (1996)'s pooled average. Ji et al. (2018) suggest at least 5 replications/imputations (their dataset has 30% missingness and around 100 participants), but discuss using more imputations for datasets with a higher percentage of missingness, smaller sample size, or complex pattern of missing data. We would therefore suggest that especially if authors have substantial amounts of missing data, they use a method robust against missing data, such as Bayesian estimation or multiple imputation (with sufficient replications).

Lag. All articles in this review use lag-1 (from time point *t* to the most recent time point previous, *t*-1), representing that they assume each timepoint is independent from others except for only the previous timepoint; other lags (such as lag-2) are more complex and require more repeated measurements (Jordan et al., 2020). However, some researchers argue that although defaulting to lag one (due to convenience or possibly for theoretical reasons) can be acceptable, ignoring higher-order lagged relationships that are present in the data can bias results (Jacobson et al., 2019). For example, if data is collected daily but variables are primarily predicting one another on a weekly basis, a lag of one (e.g., daily) will miss predictions at a higher-level lag (e.g., lag of 7 or weekly). However, a few possibilities do exist for researchers to investigate the optimal lag for their data: if a researcher is using mlVAR models, they can directly compare how different lags affect model fit using the *mlVARcompare* function and the *compareToLags* argument of the *mlVAR* R package (Epskamp et al., 2019). If a researcher is using another model or wants a more general method, they could use the Differential Time-Varying Effect Model tool (DTVEM package in R; Jacobson et al., 2019), which identifies the optimal time lag(s) in a dataset. We recommend that if researchers do investigate the effect of different lags, they report the results of the different resulting networks as sensitivity analyses.

Data Modeling

Vector Autoregressive Models (VAR). In general, VAR models regress a variable at time *t* on that same variable at time *t-1*. VAR models can also be extended to a multilevel framework where individuals are allowed to vary, called multilevel VAR (mlVAR); these are the most common type amongst reviewed articles (estimated by 95%). mlVAR models can be easily estimated using separate univariate regression equations (as generally implemented in the mlVAR package); the vast majority of authors used this estimation method (e.g., Contreras et al., 2019; Curtiss et al., 2019; Kuranova et al., 2021). Researchers generally choose to estimate

mlVAR through separate univariate regression equations because they are convenient and fast to estimate in R (while still able to estimate correlated random effects), especially with fewer than 8 nodes (Epskamp et al., 2018). However, this estimation method has weaknesses, including that it cannot directly estimate correlations between error terms or be combined with regularizations techniques to generate sparse networks (de Vos et al., 2017; Epskamp et al., 2018).

Another possibility is to use multivariate estimation through Bayesian methods. Although this is not currently available through a package in open source software (but can be done using the DSEM package in MPlus), it can model dependencies between error terms and is also more robust against missing data. Two reviewed articles used this method: Groen et al. (2020) and van Roekel et al. (2019). Groen et al. (2020) specify that they used Bayesian estimation because it can directly estimate a hierarchical model and subject-specific parameters (instead of indirectly as when a multilevel VAR model is estimated with separate univariate regression equations), can generate credible intervals (easily allowing an assessment of uncertainty or stability of parameters), and can handle missing data and unequal time observations well.

There is also the possibility to estimate a VAR model that is not multilevel, by pooling the data. This model assumes homogeneity across individuals and only estimates fixed effects (so overall is not very realistic for most data). However, this method can be used with regularization techniques and is also able to handle more nodes since there are less estimated parameters (with no random effects). Both McCuish et al. (2021) and de Vos et al. (2017) state they chose to use a pooled VAR for its regularization possibilities, since they wanted a sparse network with minimal spurious edges. However, de Vos and colleagues (2017) directly compare a pooled VAR model with a univariate estimation of a multilevel VAR model and conclude that since their data includes substantial individual differences (as is commonly the case), the multilevel model better fits their data.

There are other types of recently developed VAR approaches that are currently less common but offer great promise, although no reviewed articles utilized these approaches. For example, time-varying VAR models specifically account for time-varying processes and therefore are more robust against violations of stationarity (Haslbeck et al., 2020). For networks based on panel data or data that has less than 20 timepoints, panelVAR is a good option (Epskamp, 2020).

Other (non-VAR) Models. Although VAR models are the most common among reviewed articles, accounting for 95% of estimated models, there are also other models suited to estimate temporal networks. For example, group iterative multiple model estimation (GIMME) takes individual structures into account while estimating a group level effect (similar to multilevel modeling), but does so using an iterative search process that tries to identify associations occurring consistently throughout individuals (Lane & Gates, 2017). Among reviewed articles, only Ellison et al. (2020) estimated a GIMME model, and they specify that they chose to do so because they wanted to base their models on within-person processes and build on within-person commonalities to form group networks, instead of assuming homogeneity across persons by directly modeling with a group or hierarchical approach. Another type of model, used only by Huckins et al. (2020) among reviewed articles, is the Peter and Clark Momentary Conditional Independence (PCMCI) algorithm that is more robust against high multicollinearity and missing data while also retaining high power even for networks with many nodes (Runge et al., 2019). Although not used by any reviewed articles, sparse time series chain graphical models are another interesting alternative, as they can estimate non-Gaussian networks, model network with mixed categorical and continuous data, and can also be used with LASSO regularization (Abegaz & Wit, 2013; Jordan et al., 2020). A more in-depth overview of most of the discussed models, as well as the software/packages to estimate them, are given by both

Epskamp et al. (2018) and Jordan et al. (2020). In addition, all of the above models include discrete intervals (e.g., lags), but there is also the possibility to use continuous time dynamic models; these can be especially useful if including variables that operate at different timescales (for an overview, see Bringmann et al., 2022).

Complementary Analyses. Reviewed articles did not just estimate temporal network models, but also typically examined specific parts of the networks or compared networks. More specifically, 58% of articles compared (visually or statistically) at least two temporal networks; to statistically compare network topology or centrality indices, most articles used permutation tests. In other articles, researchers examine the structure of networks, such as looking at the edges as well as the overall network connectivity. It is essential for those examining network connectivity to decide beforehand (and ideally, preregister) a specific definition or index of connectivity. In many articles, researchers also examine how central, or connected, specific nodes are. Strength is the most commonly used centrality index within reviewed articles (in 87% of articles), and it is also recommended as the centrality index that makes the most theoretical sense to use within psychology networks (Bringmann et al., 2019). We therefore recommend using (in- and out-) strength to assess centrality if edges all have the same signs (positive or negative), or (in- and out-) expected influence if there are both negative and positive edges (Robinaugh et al., 2016). It is important to note, however, that statisticians caution researchers regarding interpreting centrality: although examining which nodes are especially connected to other nodes can help generate hypotheses for future research, a node having high centrality does not necessarily mean that it plays a causal role in the network or that it is linked with clinical outcomes (Bringmann et al., 2019).

Assessing Robustness and Accuracy. Especially since temporal networks are a new (but growing) field, it is critical to assess whether the network results are stable and accurate; this is a

concern that is currently being addressed for cross-sectional networks (e.g., Epskamp, Borsboom, et al., 2018). Few of the reviewed articles performed procedures to assess the robustness of models. However, one key method across analysis types to assess the accuracy of an estimated parameter (e.g., edge estimates) is to use bootstrapped confidence intervals, and these are very computationally intensive for multilevel models (and might not even be possible on a standard computer; Bringmann et al., 2013). It is possible to compute bootstrapped confidence intervals on pooled VAR models (e.g., with the *bootnet* package in R, as done by Lazarus et al., 2020; McCuish et al., 2021), although multilevel models are preferred to pooled models unless there is little heterogeneity among subjects (de Vos et al., 2017). It is more feasible to assess the stability of temporal networks through methods such as case-dropping (e.g., can assess if network parameters such as centrality indices or edge estimates are similar after dropping a randomly-selected percentage of participants), which can be conducted on any type of model. Jongeneel et al. (2020) employ this approach, dropping a randomly-selected 80% of participants. Epskamp (2020) also suggests that researchers can assess stability by dropping blocks of data (so as to take temporal dependencies into account, instead of dropping participants/individual rows as done to assess stability in cross-sectional networks; Epskamp et al., 2017). Another possibility is to use Bayesian estimation, as authors can then examine the credible intervals of parameters and thereby evaluate the uncertainty of estimates, as done by Groen et al. (2020). Assessing stability of temporal network analyses is important on its own merits, but it is even more crucial when there are violations of assumptions, since how these violations affect the resulting network is not yet clear (Epskamp, 2020)

Another crucial point is to examine the replicability of temporal networks. Among reviewed articles, 5% of articles split their sample to cross-validate their network models, an encouraging start to examining the replicability of temporal networks. However, that number is

still quite low, and there were no studies replicating other temporal networks, which Guloksuz et al. (2017) point out is a crucial step to assess the stability and reproducibility of network models.

Another method to examine the robustness of the analyses is to conduct a sensitivity analysis, following 33% of reviewed articles. Sensitivity analyses involve running analyses with and without a specific change (e.g., preprocessing choices like the method used to compute variables or transformations to correct for an assumption violation, or model choices) to examine its impact on the results. Sensitivity analyses can be particularly useful when researchers have doubts about the impact of a specific arbitrary analysis choice. Since the field of temporal network analyses is so new and the impact of many analysis decisions on the results is not yet fully understood, it could also be interesting for researchers to perform multiverse analyses (Steegen et al., 2016) to examine the impact of many arbitrary analysis choices at once; a limited version was done by (de Vos et al., 2017).

Reporting

We urge researchers to share their code and anonymized data if at all possible on a platform such as the Open Science Framework (OSF) or GitHub. Sharing analysis code allows others to fully understand and replicate all preprocessing and analytical choices. Sharing anonymized data might not always be possible, but it can greatly contribute to the field when possible. For one thing, temporal network analyses rely on intensive longitudinal data, and this data is extremely difficult to collect; so sharing data allows others to investigate the same dataset as well, perhaps with updated methods as the field continues to advance (Bringmann et al., 2022) or to assess the impact of arbitrary analytical choices (e.g., with a multiverse analysis; Steegen et al., 2016).

It is also important that researchers report all relevant details of their data collection, preprocessing, and analysis, which was not always the case in the reviewed articles. A lack of

clarity in reporting data collection and methodology is not uncommon in EMA research (Vachon et al., 2019), but since this field involves not only EMA research but also novel analytical methods, it is especially important to clearly report all aspects of data collection and analysis methodology so that the field can grow. We hope that researchers will find Figure 6 helpful in reviewing all essential information that should be reported in their manuscript (or supplementary materials), and we have also provided this information in a fillable checklist (https://osf.io/e6wp3/). For researchers conducting EMA data collection, they can also look at the reporting checklist by Trull & Ebner-Priemer (2020).

Limitations

This scoping review has some limitations. Since we wanted to survey a wide variety of methodology and analysis practices, we chose to limit the scope of our review to group-level temporal networks, which had to include time-series data. We therefore did not include articles focusing solely on idiographic temporal networks or panelVAR networks (based on cross-lagged panel data), although we do believe audit of these specific network types through systematic scoping reviews would be useful in the future. Our inclusion criteria did lead to some articles that were ambiguous on whether or not they should be included in the review. For example, Wichers (2014) focused on illustrating the temporal network approach, without the goal of providing an empirical contribution to the literature (other examples are discussed in the Article Selection portion of the Supplementary Materials). Another potential limitation is that we use the findings of this review to formulate recommendations, although the field is still young; just because a number of papers use a specific practice does not mean that it is optimal. However, this is why we developed our recommendation section in light of relevant literature on EMA and temporal networks, as well as the results of the scoping review.

Future of the Field

Through this scoping review, we also identify a few key areas of development for the field. First: assessing the robustness of temporal network analyses. Although there are some available methods to examine the overall stability of findings, such as through participant-dropping bootstrapping (e.g., removing a certain percentage of randomly-chosen participants and seeing how much this affects results; see Jongeneel et al., 2020), it is difficult to estimate bootstrap confidence intervals around edge/centrality estimates for multilevel temporal models (Bringmann et al., 2013). Although it is possible to pool data and not use multilevel models, this is only recommended when there is little heterogeneity among participants, since it is assumed that all participants are replications of one another (de Vos et al., 2017). One possibility to examine the robustness of point estimates is when using Bayesian estimation methods, since you can then build credible intervals around the estimated parameter; however, easily (e.g., with already built packages) using Bayesian methods to estimate temporal network requires using MPlus, a proprietary software. Being able to examine the robustness of network analyses has been deemed crucially important to the future of the field (Guloksuz et al., 2017), and we hope methods to easily assess the robustness of temporal networks will be forthcoming as well.

As we discussed in the 'Normality' section above, very little is currently known about how violations of normality affect temporal networks as well as how transformations of non-normal time-series impact results. We therefore reiterate the message from Epskamp, Waldorp, and colleagues (2018): We urge future researchers to investigate normality, assess how its violations and transformations impact temporal networks, and potentially develop new methods and models to address these challenges.

There have also been a variety of new models developed that address some of the shortcomings of current temporal networks. For example, time-varying VAR allows researchers to explicitly model non-stationarity data (Haslbeck et al., 2020), instead of correcting the data or

ignoring the assumption violation, both of which have unknown impacts on the results (Jordan et al., 2020). However, this review has not identified much use of these models. We hope that future research will continue to explore and use new models as the field grows.

Declarations

Funding

M. Annelise Blanchard and Alexandre Heeren were supported by a Coordinated Research Grant ("BParent") from the French Community of Belgium (ARC Grant 19/24-100). Alexandre Heeren (FRS-FNRS Research Associate) and M. Annelise Blanchard (Research Fellow) were both also funded by the FRS-FNRS Belgian National Science Foundation. Alba Contreras was supported by an FSR MOVE-IN Postdoctoral fellowship (UCLouvain) and World Excellence Postdoctoral fellowship from the Wallonie-Bruxelles International (WBI). These funds did not exert any influence or censorship of any kind on the present work.

Conflicts of Interest

Alexandre Heeren receives honoraria for his editorial work from Elsevier. The other authors have no known conflict of interest to disclose.

Ethics Approval, Consent to Participate and Consent for Publication

Not applicable as this is a review article and there were not human or animal subjects

CRediT Author Statement:

M. Annelise Blanchard: Conceptualization, Data curation, Formal Analysis, Investigation, Visualization, Project Administration, Writing – original draft, Funding Acquisition. **Alba Contreras:** Conceptualization, Data curation, Investigation, Verification, Writing – review & editing. **Rana Begum Kalkan:** Data curation, Investigation, Verification, Writing – review & editing. **Alexandre Heeren:** Conceptualization, Supervision, Verification, Funding Acquisition, Writing – review & editing.

Open Practice Statement

The data, materials, and R code for this scoping review are available at: https://osf.io/8647n/. This scoping review was preregistered (https://osf.io/7b45y).

References of all articles in systematic review:

(Aalbers et al., 2019; Bos et al., 2017; Bringmann et al., 2013, 2015, 2016; Contreras et al., 2020; Curtiss et al., 2019; de Vos et al., 2017; Drukker et al., 2020; Ellison et al., 2020; Elovainio et al., 2020; Faelens et al., 2021; Geraets et al., 2020; Greene et al., 2018, 2020; Groen et al., 2019, 2020; Hasmi et al., 2017; Hoffart & Johnson, 2020; Hoorelbeke et al., 2019; Huckins et al., 2020; Johnson & Hoffart, 2018; Jongeneel et al., 2020; Kaiser & Laireiter, 2019; Klippel et al., 2018; Kreiter et al., 2021; Kuranova et al., 2021; Lazarus et al., 2020; Levinson et al., 2018, 2020; Lutz et al., 2018; Martín-Brufau et al., 2020; McCuish et al., 2021; Meng et al., 2020; Oreel et al., 2019; Pannicke et al., 2021; Pavani et al., 2017; Pe et al., 2015; Rath et al., 2019; Snippe et al., 2017; Thonon et al., 2021; van Roekel et al., 2019; Wigman et al., 2015)

References

References marked with an asterisk indicate studies included in the scoping review.

* Aalbers, G., McNally, R. J., Heeren, A., de Wit, S., & Fried, E. I. (2019). Social media and depression symptoms: A network perspective. *Journal of Experimental Psychology: General*, *148*(8), 1454–1462. https://doi.org/10.1037/xge0000528

Abdullah, N., Kamaruddin, M. A., Goh, Y.-X., Othman, R., Dauni, A., Jalal, N. A., Yusuf, N. A. M., Kamat, S. A., Basri, N. H., & Jamal, R. (2021). Participants attrition in a longitudinal study: The Malaysian Cohort Study experience. *International Journal of Environmental Research and Public Health*, *18*(14), 7216. https://doi.org/10.3390/ijerph18147216

Abegaz, F., & Wit, E. (2013). Sparse time series chain graphical models for reconstructing genetic networks. *Biostatistics*, *14*(3), 586–599. https://doi.org/10.1093/biostatistics/kxt005

Barabási, A. L. (2012). The network takeover. *Nature Physics*, *8*(1), 14–16. https://doi.org/10.1038/nphys2188

Blanchard, M. A., & Heeren, A. (2022). Ongoing and future challenges of the network approach to psychopathology: From theoretical conjectures to clinical translations. In *Comprehensive Clinical Psychology—2nd Edition*. Elsevier. https://doi.org/10.1016/B978-0-12-818697-8.00044-3

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, O. J. (2013). Network analysis: An integrative approach to the structure of psychopathology. *Annual Review of Clinical Psychology*, *9*, 91–121. https://doi.org/10.1146/annurev-clinpsy-050212-185608

Borsboom, D., Deserno, M. K., Rhemtulla, M., Epskamp, S., Fried, E. I., McNally, R. J., Robinaugh, D. J., Perugini, M., Dalege, J., Costantini, G., Isvoranu, A.-M., Wysocki, A. C., van Borkulo, C. D., van Bork, R., & Waldorp, L. J. (2021). Network analysis of multivariate data in psychological science. *Nature Reviews Methods Primers*, *1*(1), 1–18. https://doi.org/10.1038/s43586-021-00055-w

* Bos, F. M., Snippe, E., Vos, S. de, Hartmann, J. A., Simons, C. J. P., Krieke, L. van der, Jonge, P. de, & Wichers, M. (2017). Can we jump from cross-sectional to dynamic interpretations of networks implications for the network perspective in psychiatry? Implications for the network perspective in psychiatry. *Psychotherapy and Psychosomatics*, *86*(3), 175–177. https://doi.org/10.1159/000453583

Bringmann, L. F., Albers, C., Bockting, C., Borsboom, D., Ceulemans, E., Cramer, A., Epskamp, S., Eronen, M. I., Hamaker, E., Kuppens, P., Lutz, W., McNally, R. J., Molenaar, P., Tio, P., Voelkle, M. C., & Wichers, M. (2022). Psychopathological networks: Theory, methods and practice. *Behaviour Research and Therapy*, *149*, 104011. https://doi.org/10.1016/j.brat.2021.104011

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

* Bringmann, L. F., Lemmens, L. H. J. M., 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. https://doi.org/10.1017/S0033291714001809

* Bringmann, L. F., Pe, M. L., Vissers, N., Ceulemans, E., Borsboom, D., Vanpaemel, W., Tuerlinckx, F., & Kuppens, P. (2016). Assessing temporal emotion dynamics using networks. *Assessment*, *23*(4), 425–435. https://doi.org/10.1177/1073191116645909

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

Cham, H., Reshetnyak, E., Rosenfeld, B., & Breitbart, W. (2017). Full information maximum likelihood estimation for latent variable interactions with incomplete indicators. *Multivariate Behavioral Research*, *52*(1), 12–30. https://doi.org/10.1080/00273171.2016.1245600

Contreras, A., Nieto, I., Valiente, C., Espinosa, R., & Vazquez, C. (2019). The study of psychopathology from the network analysis perspective: A systematic review. *Psychotherapy and Psychosomatics*, *88*(2), 71–83. https://doi.org/10.1159/000497425

* Contreras, A., Valiente, C., Heeren, A., & Bentall, R. (2020). A temporal network approach to paranoia: A pilot study. *Frontiers in Psychology*, *11*, 544565. https://doi.org/10.3389/fpsyg.2020.544565

Cramer, A. O. J., van der Sluis, S., Noordhof, A., Wichers, M., Geschwind, N., Aggen, S. H., Kendler, K. S., & Borsboom, D. (2012). Dimensions of normal personality as networks in search of equilibrium: You can't like parties if you don't like people. *European Journal of Personality*, *26*(4), 414–431. https://doi.org/10.1002/per.1866

* Curtiss, J., Fulford, D., Hofmann, S. G., & Gershon, A. (2019). Network dynamics of positive and negative affect in bipolar disorder. *Journal of Affective Disorders*, *249*(January), 270–277. https://doi.org/10.1016/j.jad.2019.02.017

Dalege, J., Borsboom, D., van Harreveld, F., van den Berg, H., Conner, M., & van der Maas, H. L. J. (2016). Toward a formalized account of attitudes: The Causal Attitude Network (CAN) model. *Psychological Review*, *123*(1), 2–22. http://dx.doi.org/10.1037/a0039802

* de Vos, S., Wardenaar, K. J., Bos, E. H., Wit, E. C., Bouwmans, M. E. J., & De Jonge, P. (2017). An investigation of emotion dynamics in major depressive disorder patients and healthy persons using sparse longitudinal networks. *PLoS ONE*, *12*(6), 1–18. https://doi.org/10.1371/journal.pone.0178586

Dejonckheere, E., & Erbas, J. (2021). Designing an experience sampling study. In *The open handbook of Experience Sampling Methodology: A step-by-step guide to designing, conducting, and analyzing ESM studies* (pp. 33–70). Center for Research on Experience Sampling and Ambulatory Methods. https://www.kuleuven.be/samenwerking/real/real-book

* Drukker, M., Peters, J. C. H., Vork, L., Mujagic, Z., Rutten, B. P. F., van Os, J., Masclee, A. A. M., Kruimel, J. W., & Leue, C. (2020). Network approach of mood and functional gastrointestinal symptom dynamics in relation to childhood trauma in patients with irritable bowel syndrome and comorbid panic disorder. *Journal of Psychosomatic Research*, *139*, 110261. https://doi.org/10.1016/j.jpsychores.2020.110261

Ellison, W. D., Levy, K. N., Newman, M. G., Pincus, A. L., Wilson, S. J., & Molenaar, P. C. M. (2020). Dynamics among borderline personality and anxiety features in psychotherapy outpatients: An exploration of nomothetic and idiographic patterns. *Personality Disorders: Theory, Research, and Treatment*, *11*(2), 131–140. https://doi.org/10.1037/per0000363

* Elovainio, M., Kuula, L., Halonen, R., & Pesonen, A.-K. (2020). Dynamic fluctuations of emotional states in adolescents with delayed sleep phase—A longitudinal network

modeling approach. *Journal of Affective Disorders*, *276*, 467–475. https://doi.org/10.1016/j.jad.2020.07.050

Epskamp, S. (2020). Psychometric network models from time-series and panel data. *Psychometrika*. https://doi.org/10.1007/s11336-020-09697-3

Epskamp, S., Borsboom, D., & Fried, E. I. (2017). Estimating psychological networks and their accuracy: A tutorial paper. *Behavior Research Methods*, *50*(1), 195–212. https://doi.org/10.3758/s13428-017-0862-1

Epskamp, S., Deserno, M. K., & Bringmann, L. F. (2019). *MlVAR: Multi-level vector autoregression*. https://CRAN.R-project.org/package=mlVAR

Epskamp, S., Rhemtulla, M., & Borsboom, D. (2017). Generalized network psychometrics: Combining network and latent variable models. *Psychometrika*, *82*(4), 904–927. https://doi.org/10.1007/s11336-017-9557-x

Epskamp, S., Waldorp, L. J., Mõttus, R., & Borsboom, D. (2018). The Gaussian Graphical Model in cross-sectional and time-series data. *Multivariate Behavioral Research*, *53*(4), 453–480. https://doi.org/10.1080/00273171.2018.1454823

* Faelens, L., Hoorelbeke, K., Soenens, B., Van Gaeveren, K., De Marez, L., De Raedt, R., & Koster, E. H. W. (2021). Social media use and well-being: A prospective experience-sampling study. *Computers in Human Behavior*, *114*, 106510. https://doi.org/10.1016/j.chb.2020.106510

* Geraets, C. N. W., Snippe, E., van Beilen, M., Pot-Kolder, R. M. C. A., Wichers, M., van der Gaag, M., & Veling, W. (2020). Virtual reality based cognitive behavioral therapy for paranoia: Effects on mental states and the dynamics among them. *Schizophrenia Research*, *222*, 227–234. https://doi.org/10.1016/j.schres.2020.05.047

* Greene, T., Gelkopf, M., Epskamp, S., & Fried, E. (2018). Dynamic networks of PTSD symptoms during conflict. *Psychological Medicine*, *48*(14), 2409–2417. https://doi.org/10.1017/S0033291718000351

* Greene, T., Gelkopf, M., Fried, E. I., Robinaugh, D. J., & Lapid Pickman, L. (2020). Dynamic network analysis of negative emotions and *DSM-5* posttraumatic stress disorder symptom clusters during conflict. *Journal of Traumatic Stress*, *33*(1), 72–83. https://doi.org/10.1002/jts.22433

* Groen, R. N., Ryan, O., Wigman, J. T. W., Riese, H., Penninx, B. W. J. H., Giltay, E. J., Wichers, M., & Hartman, C. A. (2020). Comorbidity between depression and anxiety: Assessing the role of bridge mental states in dynamic psychological networks. *BMC Medicine*, *18*(1), 308. https://doi.org/10.1186/s12916-020-01738-z

* Groen, R. N., Snippe, E., Bringmann, L. F., Simons, C. J. P., Hartmann, J. A., Bos, E. H., & Wichers, M. (2019). Capturing the risk of persisting depressive symptoms: A dynamic network investigation of patients' daily symptom experiences. *Psychiatry Research*, *271*, 640–648. https://doi.org/10.1016/j.psychres.2018.12.054

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. https://doi.org/10.1017/S0033291717001350

Hamaker, E. L., Asparouhov, T., Brose, A., Schmiedek, F., & Muthén, B. (2018). At the frontiers of modeling intensive longitudinal data: Dynamic structural equation models for the affective measurements from the COGITO study. *Multivariate Behavioral Research*, *53*(6), 820–841. https://doi.org/10.1080/00273171.2018.1446819

Haslbeck, J. M. B., Bringmann, L. F., & Waldorp, L. J. (2020). A tutorial on estimating time-varying vector autoregressive models. *Multivariate Behavioral Research*, *0*(0), 1–30. https://doi.org/10.1080/00273171.2020.1743630

* Hasmi, L., Drukker, M., Guloksuz, S., Menne-Lothmann, C., Decoster, J., van Winkel, R., Collip, D., Delespaul, P., de Hert, M., Derom, C., Thiery, E., Jacobs, N., Rutten, B. P. F., Wichers, M., & van Os, J. (2017). Network approach to understanding emotion dynamics in relation to childhood trauma and genetic liability to psychopathology: Replication of a prospective experience sampling analysis. *Frontiers in Psychology*, *8*, 1908. https://doi.org/10.3389/fpsyg.2017.01908

* Hoffart, A., & Johnson, S. U. (2020). Within-person networks of clinical features of social anxiety disorder during cognitive and interpersonal therapy. *Journal of Anxiety Disorders*, *76*, 102312. https://doi.org/10.1016/j.janxdis.2020.102312

* Hoorelbeke, K., Van den Bergh, N., Wichers, M., & Koster, E. H. W. (2019). Between vulnerability and resilience: A network analysis of fluctuations in cognitive risk and protective factors following remission from depression. *Behaviour Research and Therapy*, *116*(August 2018), 1–9. https://doi.org/10.1016/j.brat.2019.01.007

* Huckins, J. F., DaSilva, A. W., Hedlund, E. L., Murphy, E. I., Rogers, C., Wang, W., Obuchi, M., Holtzheimer, P. E., Wagner, D. D., & Campbell, A. T. (2020). Causal factors of anxiety and depression in college students: Longitudinal ecological momentary assessment and causal analysis using Peter and Clark Momentary Conditional Independence. *JMIR Mental Health*, *7*(6), e16684. https://doi.org/10.2196/16684

Isvoranu, A.-M., & Epskamp, S. (2021). *Which Estimation Method to Choose in Network Psychometrics? Deriving Guidelines for Applied Researchers*. PsyArXiv. https://doi.org/10.31234/osf.io/mbycn

Jacobson, N. C., Chow, S.-M., & Newman, M. G. (2019). The Differential Time-Varying Effect Model (DTVEM): A tool for diagnosing and modeling time lags in intensive longitudinal data. *Behavior Research Methods*, *51*(1), 295–315. https://doi.org/10.3758/s13428-018-1101-0

Ji, L., Chow, S.-M., Schermerhorn, A. C., Jacobson, N. C., & Cummings, E. M. (2018). Handling missing data in the modeling of intensive longitudinal data. *Structural Equation Modeling: A Multidisciplinary Journal*, *25*(5), 715–736. https://doi.org/10.1080/10705511.2017.1417046

* Johnson, S. U., & Hoffart, A. (2018). Metacognitive therapy versus cognitive behavioral therapy A network approach. *Frontiers in Psychology*, *9*, 2382. https://doi.org/10.3389/fpsyg.2018.02382

Jones, P. J., Heeren, A., & McNally, R. J. (2017). Commentary: A network theory of mental disorders. *Frontiers in Psychology*, *8*, 1305. https://doi.org/10.1002/wps.20375

* Jongeneel, A., Aalbers, G., Bell, I., Fried, E. I., Delespaul, P., Riper, H., van der Gaag, M., & van den Berg, D. (2020). A time-series network approach to auditory verbal hallucinations: Examining dynamic interactions using experience sampling methodology. *Schizophrenia Research*, *215*, 148–156. https://doi.org/10.1016/j.schres.2019.10.055

Jordan, D. G., Winer, E. S., & Salem, T. (2020). The current status of temporal network analysis for clinical science: Considerations as the paradigm shifts? *Journal of Clinical Psychology*. https://doi.org/10.1002/jclp.22957

* Kaiser, T., & Laireiter, A.-R. (2019). Daily dynamic assessment and modelling of intersession processes in ambulatory psychotherapy: A proof of concept study. *Psychotherapy Research*, *29*(8), 1062–1073. https://doi.org/10.1080/10503307.2018.1497213

Kjellberg, S., Haider, J., & Sundin, O. (2016). Researchers' use of social network sites: A scoping review. *Library & Information Science Research*, *38*(3), 224–234. https://doi.org/10.1016/j.lisr.2016.08.008

* Klippel, A., Viechtbauer, W., Reininghaus, U., Wigman, J., van Borkulo, C., Myin-Germeys, I., & Wichers, M. (2018). The cascade of stress: A network approach to explore differential dynamics in populations varying in risk for psychosis. *Schizophrenia Bulletin*, *44*(2), 328–337. https://doi.org/10.1093/schbul/sbx037

Kohl, C., McIntosh, E. J., Unger, S., Haddaway, N. R., Kecke, S., Schiemann, J., & Wilhelm, R. (2018). Online tools supporting the conduct and reporting of systematic reviews and systematic maps: A case study on CADIMA and review of existing tools. *Environmental Evidence*, *7*(1), 8. https://doi.org/10.1186/s13750-018-0115-5

* Kreiter, D., Drukker, M., Mujagic, Z., Vork, L., Rutten, B. P. F., van Os, J., Masclee, A. A. M., Kruimel, J. W., & Leue, C. (2021). Symptom-network dynamics in irritable bowel syndrome with comorbid panic disorder using electronic momentary assessment: A randomized controlled trial of escitalopram vs. placebo. *Journal of Psychosomatic Research*, *141*, 110351. https://doi.org/10.1016/j.jpsychores.2020.110351

* Kuranova, A., Wigman, J. T. W., Menne-Lothmann, C., Decoster, J., van Winkel, R., Delespaul, P., Drukker, M., de Hert, M., Derom, C., Thiery, E., Rutten, B. P. F., Jacobs, N., van Os, J., Oldehinkel, A. J., Booij, S. H., & Wichers, M. (2021). Network dynamics of momentary affect states and future course of psychopathology in adolescents. *PLOS ONE*, *16*(3), e0247458. https://doi.org/10.1371/journal.pone.0247458

Kwiatkowski, D., Phillips, P. C. B., Schmidt, P., & Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root. *Journal of Econometrics*, *54*(1–3), 159–178. https://doi.org/10.1016/0304-4076(92)90104-Y

Lafit, G., Adolf, J. K., Dejonckheere, E., Myin-Germeys, I., Viechtbauer, W., & Ceulemans, E. (2021). Selection of the number of participants in intensive longitudinal studies: A user-friendly shiny app and tutorial for performing power analysis in multilevel regression models that account for temporal dependencies. *Advances in Methods and Practices in Psychological Science*, *4*(1), 2515245920978738. https://doi.org/10.1177/2515245920978738

Lamers, F., Hoogendoorn, A. W., Smit, J. H., van Dyck, R., Zitman, F. G., Nolen, W. A., & Penninx, B. W. (2012). Sociodemographic and psychiatric determinants of attrition in the Netherlands Study of Depression and Anxiety (NESDA). *Comprehensive Psychiatry*, *53*(1), 63–70. https://doi.org/10.1016/j.comppsych.2011.01.011

Lane, S. T., & Gates, K. M. (2017). Automated selection of robust individual-level structural equation models for time series data. *Structural Equation Modeling: A Multidisciplinary Journal*, *24*(5), 768–782. https://doi.org/10.1080/10705511.2017.1309978

* Lazarus, G., Sened, H., & Rafaeli, E. (2020). Subjectifying the personality state: Theoretical underpinnings and an empirical example. *European Journal of Personality*, *34*(6), 1017–1036. https://doi.org/10.1002/per.2278

* Levinson, C. A., Vanzhula, I. A., Smith, T. W., & Stice, E. (2020). Group and longitudinal intra-individual networks of eating disorder symptoms in adolescents and young adults at-risk for an eating disorder. *Behaviour Research and Therapy*, *135*, 103731. https://doi.org/10.1016/j.brat.2020.103731

* Levinson, C. A., Vanzhula, I., & Brosof, L. C. (2018). Longitudinal and personalized networks of eating disorder cognitions and behaviors: Targets for precision intervention a proof of concept study. *International Journal of Eating Disorders*, *51*(11), 1233–1243. https://doi.org/10.1002/eat.22952

López-Nicolás, R., López-López, J. A., Rubio-Aparicio, M., & Sánchez-Meca, J. (2021). A meta-review of transparency and reproducibility-related reporting practices in published meta-analyses on clinical psychological interventions (2000–2020). *Behavior Research Methods*. https://doi.org/10.3758/s13428-021-01644-z

* Lutz, W., Schwartz, B., Hofmann, S. G., Fisher, A. J., Husen, K., & Rubel, J. A. (2018). Using network analysis for the prediction of treatment dropout in patients with mood and anxiety disorders: A methodological proof-of-concept study. *Scientific Reports, 8*(1), 7819. https://doi.org/10.1038/s41598-018-25953-0

Lydon-Staley, D. M., Barnett, I., Satterthwaite, T. D., & Bassett, D. S. (2019). Digital phenotyping for psychiatry: Accommodating data and theory with network science methodologies. *Current Opinion in Biomedical Engineering, 9*, 8–13. https://doi.org/10.1016/j.cobme.2018.12.003

* Martín-Brufau, R., Suso-Ribera, C., & Corbalán, J. (2020). Emotion network analysis during COVID-19 quarantine - a longitudinal study. *Frontiers in Psychology, 11*, 559572. https://doi.org/10.3389/fpsyg.2020.559572

* McCuish, E., Bouchard, M., & Beauregard, E. (2021). A network-based examination of the longitudinal association between psychopathy and offending versatility. *Journal of Quantitative Criminology*, *37*(3), 693–714. https://doi.org/10.1007/s10940-020-09462-w

McNally, R. J. (2021). Network analysis of psychopathology: Controversies and challenges. *Annual Review of Clinical Psychology*, *17*(1), 31–53. https://doi.org/10.1146/annurev-clinpsy-081219-092850

* Meng, J., Wang, X., Wei, D., & Qiu, J. (2020). State loneliness is associated with emotional hypervigilance in daily life: A network analysis. *Personality and Individual Differences*, *165*, 110154. https://doi.org/10.1016/j.paid.2020.110154

Munn, Z., Peters, M. D. J., Stern, C., Tufanaru, C., McArthur, A., & Aromataris, E. (2018). Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. *BMC Medical Research Methodology*, *18*(1), 143. https://doi.org/10.1186/s12874-018-0611-x

Myin-Germeys, I., Kasanova, Z., Vaessen, T., Vachon, H., Kirtley, O., Viechtbauer, W., & Reininghaus, U. (2018). Experience sampling methodology in mental health research: New insights and technical developments. *World Psychiatry*, *17*(2), 123–132. https://doi.org/10.1002/wps.20513

Nehrkorn-Bailey, A. M., Reardon, M. S., & Hicks Patrick, J. (2018). Some methodological and analytical issues related to real-time data capture studies. *Translational Issues in Psychological Science*, *4*(4), 349–361. http://dx.doi.org.proxy.bib.ucl.ac.be/10.1037/tps0000177

* Oreel, T. H., Borsboom, D., Epskamp, S., Hartog, I. D., Netjes, J. E., Nieuwkerk, P. T., Henriques, J. P. S., Scherer-Rath, M., van Laarhoven, H. W. M., & Sprangers, M. A. G. (2019). The dynamics in health-related quality of life of patients with stable coronary artery disease were revealed: A network analysis. *Journal of Clinical Epidemiology*, *107*, 116–123. https://doi.org/10.1016/j.jclinepi.2018.11.022

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ, 372*, n71. https://doi.org/10.1136/bmj.n71

Palmier-Claus, J. E., Myin-Germeys, I., Barkus, E., Bentley, L., Udachina, A., Delespaul, P. a. E. G., Lewis, S. W., & Dunn, G. (2011). Experience sampling research in individuals with

mental illness: Reflections and guidance. *Acta Psychiatrica Scandinavica*, *123*(1), 12–20. https://doi.org/10.1111/j.1600-0447.2010.01596.x

* Pannicke, B., Kaiser, T., Reichenberger, J., & Blechert, J. (2021). Networks of stress, affect and eating behaviour: Anticipated stress coping predicts goal-congruent eating in young adults. *International Journal of Behavioral Nutrition and Physical Activity*, *18*(1), 9. https://doi.org/10.1186/s12966-020-01066-8

Paré, G., Trudel, M.-C., Jaana, M., & Kitsiou, S. (2015). Synthesizing information systems knowledge: A typology of literature reviews. *Information & Management*, *52*(2), 183–199. https://doi.org/10.1016/j.im.2014.08.008

* Pavani, J.-B., Vigouroux, S. L., Kop, J.-L., Congard, A., & Dauvier, B. (2017). A network approach to affect regulation dynamics and personality trait-induced variations: Extraversion and neuroticism moderate reciprocal influences between affect and affect regulation strategies. *European Journal of Personality*, *31*(4), 329–346. https://doi.org/10.1002/per.2109

* Pe, M. L., Kircanski, K., Thompson, R. J., Bringmann, L. F., Tuerlinckx, F., Mestdagh, M., Mata, J., Jaeggi, S. M., Buschkuehl, M., Jonides, J., Kuppens, P., & Gotlib, I. H. (2015). Emotion-network density in major depressive disorder. *Clinical Psychological Science*, *3*(2), 292–300. https://doi.org/10.1177/2167702614540645

Peters, M. D. J., Godfrey, C. M., Khalil, H., McInerney, P., Parker, D., & Soares, C. B. (2015). Guidance for conducting systematic scoping reviews. *International Journal of Evidence-Based Healthcare*, *13*(3), 141–146. https://doi.org/10.1097/XEB.0000000000000050

* Rath, D., de Beurs, D., Hallensleben, N., Spangenberg, L., Glaesmer, H., & Forkmann, T. (2019). Modelling suicide ideation from beep to beep: Application of network analysis to

ecological momentary assessment data. *Internet Interventions*, *18*. https://doi.org/10.1016/j.invent.2019.100292

Razali, N. M., & Wah, Y. B. (2011). Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests. *Journal of Statistical Modeling and Analytics*, *2*(1), 21–33.

Robinaugh, D. J., Haslbeck, J., Waldorp, L., Kossakowski, J., Fried, E. I., Millner, A., McNally, R. J., Nes, E. H. van, Scheffer, M., Kendler, K. S., & Borsboom, D. (2019). *Advancing the network theory of mental disorders: A computational model of panic disorder*. PsyArXiv. https://doi.org/10.31234/osf.io/km37w

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

Rubin, D. B. (1996). Multiple imputation after 18+ years. *Journal of the American Statistical Association*, *91*(434), 473–489. https://doi.org/10.2307/2291635

Runge, J., Nowack, P., Kretschmer, M., Flaxman, S., & Sejdinovic, D. (2019). Detecting and quantifying causal associations in large nonlinear time series datasets. *Science Advances*, *5*(11), eaau4996. https://doi.org/10.1126/sciadv.aau4996

Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, *52*(3/4), 591–611. https://doi.org/10.2307/2333709

Shevlyakov, G., & Vilchevski, N. (2002). *Robustness in data analysis: Criteria and methods*. VSP

* Snippe, E., Viechtbauer, W., Geschwind, N., Klippel, A., de Jonge, P., & Wichers, M. (2017). The impact of treatments for depression on the dynamic network structure of mental

states: Two randomized controlled trials. *Scientific Reports*, *7*(1), 46523. https://doi.org/10.1038/srep46523

Steegen, S., Tuerlinckx, F., Gelman, A., & Vanpaemel, W. (2016). Increasing transparency through a multiverse analysis. *Perspectives on Psychological Science*, *11*(5), 702–712. https://doi.org/10.1177/1745691616658637

Steinskog, D. J., Tjøstheim, D. B., & Kvamstø, N. G. (2007). A cautionary note on the use of the Kolmogorov–Smirnov test for normality. *Monthly Weather Review*, *135*(3), 1151–1157. https://doi.org/10.1175/MWR3326.1

Terluin, B., Boer, M. R. de, & Vet, H. C. W. de. (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

* Thonon, B., Contreras, A., & Larøi, F. (2021). Motivation in schizophrenia: Preliminary findings of a theory-driven approach using time-series network analysis. *Current Psychology*. https://doi.org/10.1007/s12144-020-01321-y

Tricco, A. C., Lillie, E., Zarin, W., O'Brien, K. K., Colquhoun, H., Levac, D., Moher, D., Peters, M. D. J., Horsley, T., Weeks, L., Hempel, S., Akl, E. A., Chang, C., McGowan, J., Stewart, L., Hartling, L., Aldcroft, A., Wilson, M. G., Garritty, C., … Straus, S. E. (2018). PRISMA extension for scoping reviews (PRISMA-ScR): Checklist and explanation. *Annals of Internal Medicine*, *169*(7), 467–473. https://doi.org/10.7326/M18-0850

Trull, T. J., & Ebner-Priemer, U. W. (2009). Using experience sampling methods/ecological momentary assessment (ESM/EMA) in clinical assessment and clinical research: Introduction to the special section. *Psychological Assessment*, *21*(4), 457–462. https://doi.org/10.1037/a0017653

Trull, T. J., & Ebner-Priemer, U. W. (2020). Ambulatory assessment in psychopathology research: A review of recommended reporting guidelines and current practices. *Journal of Abnormal Psychology, 129*(1), 56–63. http://dx.doi.org/10.1037/abn0000473

Vachon, H., Viechtbauer, W., Rintala, A., & Myin-Germeys, I. (2019). Compliance and retention with the experience sampling method over the continuum of severe mental disorders: Meta-analysis and recommendations. *Journal of Medical Internet Research, 21*(12), e14475. https://doi.org/10.2196/14475

Van Der Maas, H. L. J., Kan, K.-J., Marsman, M., & Stevenson, C. E. (2017). Network models for cognitive development and intelligence. *Journal of Intelligence, 5*(2), 16. https://doi.org/10.3390/jintelligence5020016

* van Roekel, E., Heininga, V. E., Vrijen, C., Snippe, E., & Oldehinkel, A. J. (2019). Reciprocal associations between positive emotions and motivation in daily life: Network analyses in anhedonic individuals and healthy controls. *Emotion, 19*(2), 292–300. http://dx.doi.org/10.1037/emo0000424

Varese, F., Haddock, G., & Palmier-Claus, J. (2019). Designing and conducting an experience sampling study. In *Experience Sampling in Mental Health Research* (pp. 8–17). Routledge.

Weston, S. J., Ritchie, S. J., Rohrer, J. M., & Przybylski, A. K. (2019). Recommendations for increasing the transparency of analysis of preexisting data sets. *Advances in Methods and Practices in Psychological Science, 2*(3), 214–227. https://doi.org/10.1177/2515245919848684

Wichers, M. (2014). The dynamic nature of depression: A new micro-level perspective of mental disorder that meets current challenges. *Psychological Medicine, 44*(1349–1360), 13.

Wichers, M., Wigman, J. T. W., Bringmann, L. F., & de Jonge, P. (2017). Mental disorders as networks: Some cautionary reflections on a promising approach. *Social Psychiatry and Psychiatric Epidemiology*, *52*(2), 143–145. https://doi.org/10.1007/s00127-016-1335-z

* Wigman, J. T. W., van Os, J., Borsboom, D., Wardenaar, K. J., Epskamp, S., Klippel, A., Viechtbauer, W., Myin-Germeys, I., & Wichers, M. (2015). Exploring the underlying structure of mental disorders: Cross-diagnostic differences and similarities from a network perspective using both a top-down and a bottom-up approach. *Psychological Medicine*, *45*(11), 2375–2387. https://doi.org/10.1017/s0033291715000331

Zarin, W., Veroniki, A. A., Nincic, V., Vafaei, A., Reynen, E., Motiwala, S. S., Antony, J., Sullivan, S. M., Rios, P., Daly, C., Ewusie, J., Petropoulou, M., Nikolakopoulou, A., Chaimani, A., Salanti, G., Straus, S. E., & Tricco, A. C. (2017). Characteristics and knowledge synthesis approach for 456 network meta-analyses: A scoping review. *BMC Medicine*, *15*(1), 3. https://doi.org/10.1186/s12916-016-0764-6