

# The Sum of All Fears: Comparing Networks Based on Symptom Sum-Scores

Jonas M. B. Haslbeck<sup>1\*</sup>, Oisín Ryan<sup>2\*</sup>, and Fabian Dablander<sup>1\*</sup>

<sup>1</sup> Department of Psychological Methods, University of Amsterdam

<sup>2</sup> Department of Methodology and Statistics, Utrecht University

## Abstract

Researchers are often interested in comparing statistical network models estimated from groups that are defined by the sum-score of the modeled variables. A prominent example is an analysis that compares networks of individuals with and without a diagnosis of a certain disorder. Recently, several authors suggested that this practice may lead to invalid inferences by introducing Berkson's bias. In this article, we show that whether bias is present or not depends on which research question one aims to answer. We review five possible research questions one may have in mind when separately estimating network models in groups that are based on sum-scores. For each research question, we provide an illustration with a simulated bivariate example and discuss the nature of the bias, if present. We show that if one is indeed interested in the network models of the groups defined by the sum-score, no bias is introduced. However, if one is interested in differences across groups defined by a variable other than the sum-score, detecting population heterogeneity, the network model in the general population, or inferring causal relations, then bias will be introduced in most situations. Finally, we discuss for each research question how bias can be avoided.

## Translational Abstract

Researchers in clinical psychology and psychiatry increasingly study relationships between symptoms of mental disorders using network models. In this context, it is often interesting to compare network models across groups, for example based on gender, age, or whether individuals are diagnosed with a disorder. The latter case is different from the first two, because diagnosis groups are themselves defined by the presence or absence of symptoms. Recently, several researchers have suggested that creating groups in this way leads to biased estimates. In this article, we show that whether or not bias occurs depends on the research question at hand. We consider five different types of research questions and in each case determine what the target of our analysis is, whether bias is present, and if so, where that bias is coming from. In each case, we outline how that bias could best be dealt with or avoided, showing that each research question requires a qualitatively different approach.

*Keywords:* Network models, Group comparison, Berkson's bias

The network approach to psychopathology conceptualizes mental disorders as systems of causally interacting symptoms (Borsboom, 2017; Borsboom & Cramer, 2013; Schmittmann et al.,

2013). From this perspective, psychopathology arises not from the presence of some latent disease variable (e.g., “depression”) causing symptoms, but from the direct interactions between the symptoms themselves. A popular strategy to gain insight into these interactions is to analyze symptom data with pairwise statistical network models such as the Gaussian graphical model (GGM) or the Ising model (e.g., Epskamp et al., 2018).

A key prediction of the network approach is that healthy and unhealthy individuals differ in the patterns of causal relations between their symptoms, for example with respect to the role particular symptoms play in the network, or how densely connected the symptoms are overall (Borsboom, 2017; Cramer et al., 2010). To evaluate this prediction, many studies have investigated differences between healthy and unhealthy individuals by comparing estimated network models across groups (e.g., Heeren & McNally, 2018; Levinson et al., 2018; Meier et al., 2019; Santos et al., 2017; Southward & Cheavens, 2018; Van Rooijen et al., 2018). To define healthy and unhealthy groups, a popular

This article was published Online First December 16, 2021.

Jonas M. B. Haslbeck  <https://orcid.org/0000-0001-9096-7837>

Oisín Ryan  <https://orcid.org/0000-0003-3698-6396>

Fabian Dablander  <https://orcid.org/0000-0003-2650-6491>

\*All authors contributed equally to this work and should be considered joint first authors.

We thank Tessa Blanken, Riet van Bork, Denny Borsboom, Laura Bringmann, Marie Deserno, Sacha Epskamp, Eiko Fried, Max Hinne, Don Robinaugh, Jill de Ron, Noémi Schuurman, and Lourens Waldorp for helpful discussions and comments on earlier versions of this article.

Correspondence concerning this article should be addressed to Jonas M. B. Haslbeck, Department of Psychological Methods, University of Amsterdam, Nieuwe Achtergracht 129-B, 1018 WT, Amsterdam, the Netherlands. Email: [jonashaslbeck@gmail.com](mailto:jonashaslbeck@gmail.com) or [www.jonashaslbeck.com](http://www.jonashaslbeck.com)

choice is to use the *DSM-5* definition of the disorder at hand, which typically relies on a cut-off value based on the symptom sum-score (e.g., at least five out of nine symptoms for major depression; American Psychiatric Association, 2013).

However, this type of analysis has been put into question by several authors who have suggested that defining groups based on the sum-score leads to biased inferences (for example, Birkeland et al., 2020; Dablander et al., 2019; De Ron et al., 2019; Epskamp & Fried, 2018; Fritz et al., 2018; Funkhouser et al., 2020; Lazarov et al., 2020; Meier et al., 2019; Mkhitarian et al., 2019; Mullarkey et al., 2021; Robinaugh et al., 2020). How could such a bias arise in principle? Bias is defined as the discrepancy between the *expectation of an estimator* (i.e., the average estimate obtained from many random samples) and a particular *target of inference*. Researchers may obtain biased estimates for a number of reasons. One way is by intentionally choosing to use a biased estimator in order to obtain estimates with some other desirable qualities (e.g., using regularization techniques to reduce variance; Hastie et al., 2015). A second way to obtain biased estimates is by unintentionally using an estimator which does not correspond to one's particular target of inference. It is this latter case that is related to conditioning on the sum-score and is the focus of this article.

While there are numerous historical treatments of biases related to conditioning on so-called endogenous variables in the statistical literature (e.g., Berkson, 1946; Muthén, 1989; Pearl, 2009), De Ron et al. (2019) first investigated this problem in the context of network model comparisons based on symptom sum-scores. They studied the specific case in which one's research question concerns the difference in network structure between two *latent* groups, but instead one investigates the differences between groups defined by the sum-score. Because the target of inference (the networks in the latent groups) does not correspond to the estimator used (the networks in the sum-score groups), this procedure necessarily leads to bias, except in the special case in which the latent groups happen to be the same as the sum-score groups.

But group comparisons can also be motivated by other types of research questions. These could focus on the groups that *are* defined by the sum-score (such as diagnosed vs. not diagnosed groups), the detection of population heterogeneity, the network model in the general population, or the estimation of causal effects. Crucially, in each of these cases the target of inference is different. Consequently, the answer to the question whether any bias is present, and if so, how to mitigate it, depends on the type of research question one aims to answer by comparing network models across groups that are defined by the sum-score.

In this article, we consider five types of research questions one may have in mind when comparing groups based on the sum-score, including the one originally studied by De Ron et al. (2019). For each type of research question, we use a simulated bivariate example as an illustration and discuss whether bias is present. We show that if one is indeed interested in the groups based on the sum-score, no bias is present. However, if one's research question concerns the networks based on another (latent) grouping variable, detecting population heterogeneity, the network model in the general population, or causal relations, biases exist. In each of these cases we pinpoint the exact nature of the bias and discuss strategies to avoid it. We thereby hope to add clarity for

researchers about when comparing groups that are defined on the sum-score is problematic and how to avoid possible biases.

### Different Research Questions, Different Biases

In this section, we consider five research questions that one may have in mind when separately estimating network models in groups defined by a symptom sum-score. To discuss possible biases that arise for each research question, we use the simplest nontrivial example: Two continuous variables  $X_1$  and  $X_2$ , with two groups defined by their sum. These two continuous variables could be interpreted as severity scores of two symptoms of a given disorder. In this example, the statistical network structure simply takes the form of a correlation parameter. However, all our conclusions extend to more than two variables, conditional dependence measures such as partial correlations, more than two groups, and the binary and mixed variable case.

### Network Models of Diagnosed and Not Diagnosed Groups

The first type of research question we consider concerns the investigation of statistical relationships between symptom variables in both diagnosed and not diagnosed groups of individuals. Because symptom sum-scores are the primary criteria used to diagnose psychopathology in the *DSM-5* (American Psychiatric Association, 2013), researchers may define groups of diagnosed and not diagnosed individuals based on a sum-score cut-off of the symptom variables themselves.<sup>1</sup> This allows researchers to investigate how patterns of symptom relationships differ across diagnosed and not diagnosed groups. In statistical terms, this type of research question concerns the statistical relationships between symptoms *conditional* on diagnostic status.

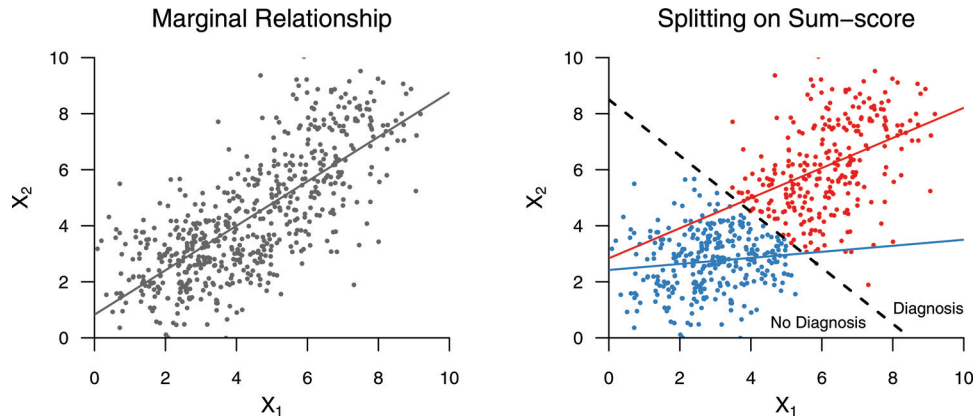
To answer this type of research question, we split the data using the sum-score of our symptoms and compute the correlation in each group. Figure 1 illustrates this process for our bivariate example.<sup>2</sup> The left panel shows the relationship between the two symptoms  $X_1$  and  $X_2$  in the overall population, and the solid line displays the best fitting regression line, which indicates a strong positive relationship ( $r = 0.74$ ). This reflects the common empirical finding that symptoms are positively correlated. We now split the sample based on the sum-score  $S = X_1 + X_2$ , defining individuals with a sum-score of  $S < 8.50$  as belonging to the no diagnosis group ( $D = 0$ ), and individuals having a sum-score of  $S \geq 8.50$  as belonging to the diagnosed group ( $D = 1$ ). The dashed diagonal line in the right panel of Figure 1 depicts this cut-off value, with the diagnosis group and no-diagnosis group indicated by the red and blue points, respectively. We can now estimate the parameters of the network models separately in each group, obtaining the

<sup>1</sup> In the present article we take symptom sum-scores to be sufficient in defining diagnosed and not diagnosed groups for the sake of simplicity. In practice, additional criteria are often used, such as in the *DSM-5* definition of depression, which requires 5 or more of 9 symptoms (including at least 1 of depressed mood and loss of interest or pleasure) [...], (American Psychiatric Association, 2013). The discussion of biases in the current article also applies to the use of criteria additional to the sum-score to create diagnosed and not diagnosed groups.

<sup>2</sup> Code to reproduce the figures is available from <https://github.com/fdabl/Sum-Score-Paper>.

**Figure 1**

Relationships Between  $X_1$  and  $X_2$  in the Entire Population and in Subgroups Defined by Their Sum-Score



Note. Left: Relationship between severity of symptoms  $X_1$  and  $X_2$ . The solid line indicates the best fitting regression line ( $r = 0.74$ ). Right: Relationship between severity of symptoms  $X_1$  and  $X_2$  in diagnosed and not diagnosed individuals. The dashed black line indicates the split of the symptom sum-score at the value 8.50; individuals with sum-score larger 8.50 are above this line (diagnosis), and subject with sum-score lower 8.50 are below this line (no diagnosis). The solid lines indicate the best fitting regression lines in the separate groups, with correlations  $r_{D=0} = 0.12$  and  $r_{D=1} = 0.40$ , respectively. See the online article for the color version of this figure.

correlation between both symptoms conditional on diagnosis. We find that in the group without diagnosis the correlation is  $r_{D=0} = 0.12$ , while the correlation in the group with diagnosis is higher  $r_{D=1} = 0.40$ .<sup>3</sup>

Given the present research question and analysis, are these correlation coefficients biased? Recall that bias is defined as a discrepancy between the expectation of the estimator and the target of inference. In the present case, the estimates are the sample correlation coefficients in the groups with ( $D = 1$ ) and without ( $D = 0$ ) diagnosis, and the targets of inference are the population correlation coefficients in the groups with ( $D = 1$ ) and without ( $D = 0$ ) diagnosis. Consequently, there is no bias.

Critically, the absence of bias in the context of this research question does not imply the absence of bias in the context of other research questions. Indeed, for the four research questions discussed in the remainder of this article, different biases arise when trying to answer them by splitting the data based on the sum-score. We discuss the bias in each context in detail and describe possible remedies that may help researchers to avoid biased inferences.

### Network Models of Groups Defined by Another Variable

The second type of research question we consider also concerns the investigation of how statistical relationships between symptoms depend on group membership, but where the grouping variable of interest ( $G$ ) is different from the diagnostic grouping variable ( $D$ ). The bias in this scenario has previously been described in the network literature by De Ron et al. (2019), who take  $G$  to represent a latent variable “healthy” versus “unhealthy” as distinct from “diagnosed” versus “not diagnosed.” In this situation, the network model conditional on  $D$  will be a biased estimate of the network model conditional on  $G$ . Figure 2 illustrates this for our bivariate example.

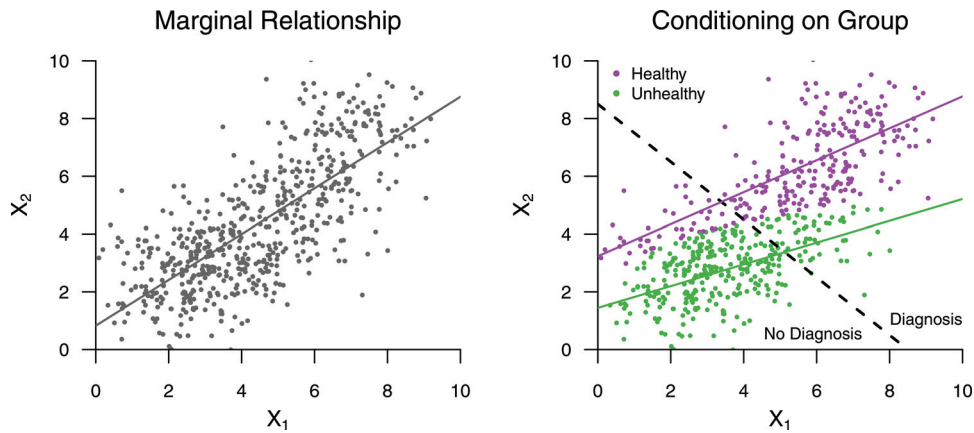
Comparing the regression lines in the two groups based on diagnosis,  $D = 1$  and  $D = 0$  (right panel of Figure 1), with the regression lines in the two groups defined by  $G$  (right panel of Figure 2) shows that they are different. Here, the sample correlation coefficients  $r_{G=0}$  and  $r_{G=1}$  are unbiased estimators of the population correlations in groups  $G = 0$  and  $G = 1$ , but, as we can see, membership of  $D$  and  $G$  do not correspond. Consequently, if we use the sample correlation in group  $D = 0$  ( $r_{D=0} = 0.12$ ) as an estimate of the population correlation in Group  $G = 0$  (which is much higher, as indicated by the sample correlation  $r_{G=0} = 0.55$ ), then this estimate will be biased.

In which situations is one dealing with such an alternative grouping variable  $G$ ? There are two types of situations. In the first one we can obtain the grouping variable from the data. For example, the grouping variable might simply be available (e.g., sex at birth) or is a function of the data that is different from the sum-score. In this situation, avoiding bias is easy: One splits the data using  $G$  in the groups of interest, and estimates the statistical network models in each of the groups.

The second situation occurs if  $G$  is not directly observable and cannot be obtained from the data. This is the situation researchers are in if they wish to study, for example, the variable “depression,” but they assume that membership of the depressed or not depressed group is not equivalent to meeting the diagnostic criteria for depressed or not depressed, as indicated by the symptom variables. Because the grouping variable  $G$  is unobserved, it cannot be used to split the data into groups, and so we need to approximate it somehow. One option to approximate  $G$  would be to use the diagnostic Group  $D$ , as we did above. The better  $D$  approximates  $G$ , the smaller the bias. But unless  $D$  and  $G$  are equivalent, some bias

<sup>3</sup> The result that the correlation is stronger in the group with diagnosis in this example reflects the finding that the symptom networks in groups with diagnosis are more dense than the symptom networks of groups without diagnosis (although the opposite result has also been found in the literature; for a review see Robinaugh et al., 2020).

**Figure 2**  
Relationships Between  $X_1$  and  $X_2$  in the Entire Population and in Subgroups Defined by Alternative Grouping Variable



*Note.* Left: Relationship between severity of symptoms  $X_1$  and  $X_2$ . The solid line indicates the best fitting regression line ( $r = 0.74$ ). Right: Relationship between severity of symptoms  $X_1$  and  $X_2$  in different groups  $G$ . The dashed black line indicates the split of the symptom sum-score at the value 8.50. The purple and green colors indicate “healthy” ( $G = 0$ ) and “unhealthy” ( $G = 1$ ) latent groups, whose correlation between the symptoms is  $r_{G=0} = 0.55$  and  $r_{G=1} = 0.71$ , respectively. See the online article for the color version of this figure.

will be present. The obvious strategy to reduce bias in this scenario is to better specify what this alternative grouping variable  $G$  really is, or how it could in principle be measured. For example, one might be able to better approximate  $G$  using other variables such as biological or environmental risk factors. However, this is only possible if  $G$  is well defined. This shows that the fundamental reason for the bias in this situation is not a statistical problem related to splitting the data into two groups, but the conceptual problem of not specifying what the latent variable  $G$  is and how it can be measured.

Let’s take a step back and consider again the specific example of  $G$  being the latent variable “depression.” Assuming that such a variable exists implies that we take the position that depression can be conceptualized along a single dimension. However, this position is at odds with the premise of the network approach as a theoretical framework, which posits that mental disorders arise from direct interactions between symptoms (Borsboom, 2017). It therefore seems unclear whether it is conceptually consistent to analyze symptom networks separately for groups that represent different levels of depression on a unidimensional scale. An alternative way to investigate how the network structure depends on symptom activation that is conceptually consistent with the network approach would be to specify a model in which each pairwise interaction is moderated by each variable in the model (De Ron et al., 2019; Haslbeck, 2020; Haslbeck et al., 2019). This way, the interactions between symptoms can be dependent on the values of the modeled variables, without assuming an underlying latent variable.

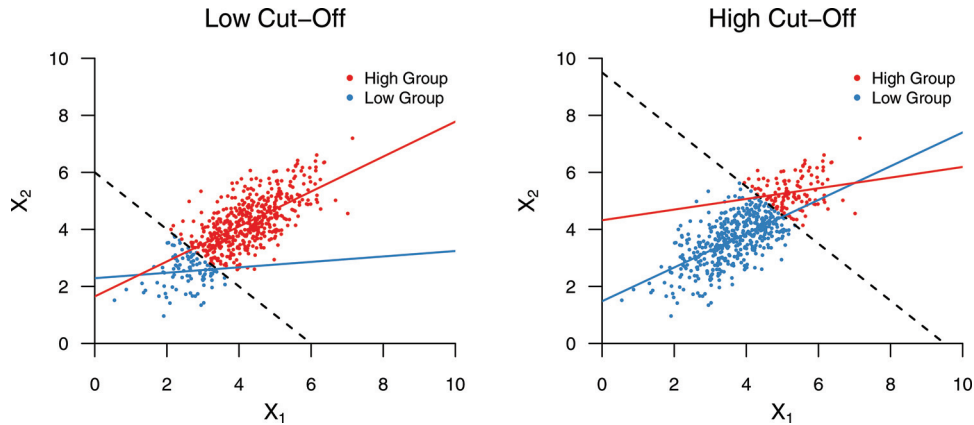
### Population Heterogeneity

The third research question concerns the detection of population heterogeneity, that is, detecting the presence of subgroups defined

by distinct statistical characteristics (such as means, variances, correlations, and/or partial correlations). This research question differs the second research question, where the presence of distinct subgroups was assumed a priori as a latent grouping variable  $G$ , and the first research question, where distinct subgroups were imposed based on sum-scores and diagnostic criteria  $D$ . We will treat the identification of population heterogeneity in statistical terms as a question about *mixtures* of distributions: Are the magnitude of the symptoms and the relations between them best described with a single multivariate distribution, or does a mixture of two or more such distributions better describe the data? In the latter case, we refer to the population as being heterogeneous, while in the former case we refer to the population as being homogeneous. This research question is conceptually similar to the question whether a latent class structure underlies the data (cf. Borsboom et al., 2016). Here, we discuss to what extent population heterogeneity can be detected by splitting the data on the sum-score.

A perhaps intuitive line of argument would be that, if the network parameters in the two groups defined on the sum-score differ, the population is heterogeneous. This reasoning is flawed, however, and Figure 3 illustrates why: In both panels we show the same data generated from a *single* bivariate Gaussian distribution, which means that there is *no* population heterogeneity. In the two panels we split the data at different cut-off values of the sum-score (depicted by the dashed diagonal lines). We see that the correlations in the “high” and “low” sum-score groups are different, even though the data were generated from a single distribution. Furthermore, by comparing the left and right panels, we see that those group differences are a function of the chosen cut-off value. In fact, the correlations in the two groups will be different for almost all cut-off values. Critically, this shows that network models will be different across groups defined on the sum-score, even if the

**Figure 3**  
Relationship Between  $X_1$ ,  $X_2$  for Different Sum-Score Cut-Off Values



*Note.* Relationships between severity of symptoms  $X_1$  and  $X_2$  in the low and high groups, for different cut-off values on the sum-score. Both panels illustrate that the network models in the two groups can be very different while the population is homogeneous. See the online article for the color version of this figure.

studied population is homogeneous. Consequently, comparing groups based on the sum-score cannot be a valid method to detect population heterogeneity.

To directly assess whether population heterogeneity exists, researchers may wish to use mixture models (e.g., McLachlan & Peel, 2000), as has been suggested previously in the network literature (e.g., Brusco et al., 2019; De Ron et al., 2019). In particular, one can fit mixture models with a varying number of mixture components or groups (e.g., 1, 2, or 3) to the data and perform model selection, for instance using information criteria such as the BIC to find the best fitting mixture model (Leroux, 1992; Steele & Raftery, 2010). This can be done in a confirmatory manner by, for example, using the sum-score as a way to define the group membership of each observation, or in an exploratory manner by learning the group membership of each observation. If model selection returns a single component, we would speak of a homogeneous population; if it returns two or more components, we would speak of a heterogeneous population.

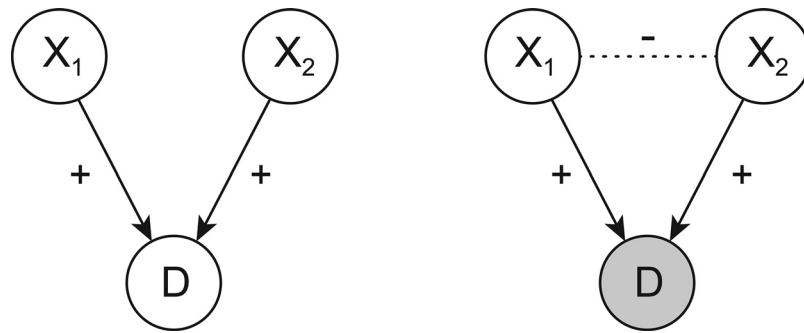
### Network Model in the General Population

The fourth type of research question we consider is about the network structure in the *general population*, which consists of both diagnosed and not diagnosed (or, if we choose to consider such a grouping variable, “healthy” and “unhealthy”) individuals. The network model in the general population is also called the *marginal network model*, because it is obtained by marginalizing (i.e., averaging) over all variables (e.g., group memberships) that may have an effect on the network. If one has a representative sample from the general population, the approach one should take to answer this type of research question is simple: Estimate the network model on the full sample, without splitting the data into groups. This provides an unbiased estimate of the population network model, to the extent that the sample is indeed representative of the general population.

The problematic case, however, is when one takes a sample from a specific subpopulation and uses it to make inferences about the general population. This reflects the situation initially studied by Berkson (1946): Suppose one wishes to study the correlation between lung cancer ( $X_1$ ) and diabetes ( $X_2$ ), and that the correlation in the general population is zero. When one looks at the correlation between these two diseases among patients in a hospital, however, then  $X_1$  and  $X_2$  are negatively associated. This is because both lung cancer and diabetes increase one’s chances of visiting the hospital: If one is in the hospital but does not have lung cancer, then it is more likely that one does have diabetes than it is for hospitalized individuals who do have lung cancer. The correlation between  $X_1$  and  $X_2$  in the hospital is a conditional correlation, since it is the correlation conditional on being in the hospital. The bias induced by using a conditional correlation as an estimate for the marginal correlation has traditionally been referred to as *Berkson’s bias* (e.g., Snoep et al., 2014).

The situation that marginal and conditional relationships need not be the same is also illustrated in the example in Figure 1. Here, the correlation in the whole sample (left panel) is different than the correlation in either of the two separate groups (right panel). Clearly, in such a situation taking the correlation of one of the groups as an estimate of the correlation in the general population would lead to bias. Notably, the converse is also true: Estimating the correlation using the whole sample would yield a biased estimate of the correlation in each group (a situation previously discussed in the network literature by Hoffman et al., 2019). However, it should be noted that researchers who purposefully split their sample according to the sum-score are unlikely to have marginal relationships in mind as their target of inference, since the marginal relationships can simply be estimated directly without splitting the sample in the first place. Splitting the sample by the sum-score group would suggest that we are interested in some kind of conditional relationship instead, and so Berkson’s bias does not occur in that case.

**Figure 4**  
Causal Graph Connecting Variables and Sum-Score



*Note.* Left: Causal graph with variables  $X_1$  and  $X_2$  having a positive causal effect on variable  $D$ . When  $D$  is defined by the sum-score of  $X_1$  and  $X_2$ , it is fully determined by  $X_1$  and  $X_2$ . Right:  $X_1$  and  $X_2$  become negatively associated when conditioning on the collider  $D$ .

## Causal Relations

The final scenario concerns the use of the conditional relationship between two variables  $X_1$  and  $X_2$  given diagnosis  $D$ , when the research question concerns the *causal* relationship between  $X_1$  and  $X_2$ . From a causal perspective  $D$  would be called a common effect or a collider of  $X_1$  and  $X_2$ , and *collider bias* refers to the fact that the conditional correlation of  $X_1$  and  $X_2$  given  $D$  is a biased estimate of the direct causal effect between  $X_1$  and  $X_2$ . Unlike the scenarios described in all previous sections, the target of inference here concerns causal relationships, not merely statistical ones (Pearl, 2009).

We illustrate the problem of collider bias using the following example. The left panel of Figure 4 depicts causal relationships between three variables,  $X_1$ ,  $X_2$ , and  $D$ , as a directed acyclic graph (DAG), which is a widespread and well-developed formalism for describing directed causal structures. In this scenario, we see that both  $X_1$  and  $X_2$  have a *positive* direct causal effect on  $D$  (that is,  $X_1 \rightarrow D$  and  $X_2 \rightarrow D$ ) but that  $X_1$  and  $X_2$  have no direct causal effect on one another. We can think about these causal effects as describing what would happen if we were to intervene on a variable: Setting  $X_1$  to a larger value would increase the value of  $D$ , but have no effect on the value of  $X_2$ . These causal relationships in turn imply a certain set of statistical relationships:  $D$  is dependent on  $X_1$  and  $X_2$  both marginally and when conditioning on  $X_2$  or  $X_1$ , respectively. However, while  $X_1$  and  $X_2$  are marginally independent of each other, conditioning on  $D$  induces a *negative* statistical dependency between them. As such, the conditional dependency between  $X_1$  and  $X_2$  given  $D$  is a biased estimate of the causal dependency between them. Note that if  $X_1$  and  $X_2$  were positively related, conditioning on  $D$  would result in a downward biased—but not necessarily negative—estimate, too.

Collider bias as we have outlined here is an issue when using statistical network models (based on sum-score splits) with the goal of inferring underlying directed causal structures. However, researchers should also be aware that statistical network models are likely to induce collider bias even in the absence of sum-score splitting, and that the use of statistical network models to infer

directed causal structures is problematic (e.g., Dablander & Hinne, 2019; Ryan et al., 2019). Researchers who wish to infer a directed causal structure in the general population from observational data—with all caveats attached—have a plethora of tools at their disposal (for an overview see Spirtes & Zhang, 2016). Causal search methods generally do not allow for heterogeneity in the causal structure across groups, however. To search for heterogeneous causal structures, researchers could instead use a “mixture of DAGs” approach, which combines mixture modeling techniques as described above with causal search procedures (e.g., Saeed et al., 2020). The royal road to estimating causal effects, however, remains performing an experiment.

## Conclusions

In this article, we reviewed five types of research questions one may have in mind when separately estimating statistical network models across groups that are defined based on the sum-score. For each type of research question, we discussed the presence of bias using a bivariate example, and our findings are summarized in Table 1.

Our analysis adds clarity for applied researchers and methodologists concerned by potential problems associated with estimating network models from groups defined by the sum-score. We showed that even though statistical network models are a relatively new addition to the statistical toolbox of psychologists, the issues related to conditioning have been identified previously in more general settings. In the cases in which the problems were of a statistical nature, there were clear solutions: If one is indeed interested in the network models conditional on the sum-score, there is no problem; if one is interested in the network model in the general population, one should not condition on any variable; and if one aims to detect population heterogeneity, one should use appropriate statistical methods to do so. However, in case one would like to estimate network models in groups defined by a latent variable, the problem is primarily conceptual: One needs to clearly characterize this latent variable and specify how to measure it. Finally, if the goal is to move beyond statistical relationships and

**Table 1**

Summary of Which Biases Are Present for Different Types of Research Goals When Analyzing Network Models Separately in Groups Defined by the Sum-Score

Research goal	Target of inference	Bias	Suggested approach
Estimating network models of diagnosed and undiagnosed groups	$r_{D=1}, r_{D=0}$	No	Use symptom sum-score to split groups.
Estimating network models of healthy and unhealthy groups (where health $\neq$ diagnosis)	$r_{G=1}, r_{G=0}$	Yes	If possible, split on grouping variable $G$ . If $G$ unobservable, solution depends on understanding $G$ . Possible that $D$ may be the best available approximation.
Detecting population heterogeneity	Number of components	Yes	(Gaussian) mixture models to test for $K = 1, 2, \dots$ components.
Estimating network model in the general population	$r$	Yes	Do not use symptom sum score to split groups. Obtain representative sample of whole population. Marginalize over all grouping variables.
Estimating causal relationships between symptoms	$E[X_2   do(X_1), D = 1],$ $E[X_2   do(X_1), D = 0]$	Yes	Use causal search procedures; causal search with mixtures of DAGs; if possible, conduct experiments

Note. The target of inference  $r$  refers to the correlation in the bivariate case used throughout the article.

to discover causal relations, one needs to use appropriate methods for causal inference. In sum, we showed that the key to avoiding biases related to conditioning on sum-scores is to clearly specify what the target of inference is. Only then can one choose an appropriate method that avoids bias.

## References

- American Psychiatric Association. (2013). *Diagnostic and Statistical manual of mental disorders: DSM-5* (5th ed.).
- Berkson, J. (1946). Limitations of the application of fourfold table analysis to hospital data. *Biometrics Bulletin*, 2(3), 47–53. <https://doi.org/10.2307/3002000>
- Birkeland, M. S., Greene, T., & Spiller, T. R. (2020). The network approach to posttraumatic stress disorder: A systematic review. *European Journal of Psychotraumatology*, 11(1), 1700614. <https://doi.org/10.1080/20008198.2019.1700614>
- Borsboom, D. (2017). A network theory of mental disorders. *World Psychiatry*, 16(1), 5–13. <https://doi.org/10.1002/wps.20375>
- Borsboom, D., & Cramer, A. O. (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., Rhemtulla, M., Cramer, A., Maas, H. L. V D., Scheffer, M., & Dolan, C. (2016). Kinds versus continua: A review of psychometric approaches to uncover the structure of psychiatric constructs. *Psychological Medicine*, 46(8), 1567–1579. <https://doi.org/10.1017/S0033291715001944>
- Brusco, M. J., Steinley, D., Hoffman, M., Davis-Stober, C., & Wasserman, S. (2019). On Ising models and algorithms for the construction of symptom networks in psychopathological research. *Psychological Methods*, 24(6), 735–753. <https://doi.org/10.1037/met0000207>
- Cramer, A. O., Waldorp, L. J., Van Der Maas, H. L., & Borsboom, D. (2010). Comorbidity: A network perspective. *Behavioral and Brain Sciences*, 33(2-3), 137–150. <https://doi.org/10.1017/S0140525X09991567>
- Dablander, F., & Hinne, M. (2019). Node centrality measures are a poor substitute for causal inference. *Scientific Reports*, 9(1), 1–13. <https://doi.org/10.1038/s41598-019-43033-9>
- Dablander, F., Epskamp, S., & Haslbeck, J. M. (2019). Studying statistics anxiety requires sound statistics: A comment on Siew, McCartney, and Vitevitch (2019). *Scholarship of Teaching and Learning in Psychology*, 5(4), 319–323. <https://doi.org/10.1037/stl0000159>
- De Ron, J., Fried, E. I., & Epskamp, S. (2019). Psychological networks in clinical populations: Investigating the consequences of Berkson's bias. *Psychological Medicine*, 51(1), 168–176.
- Epskamp, S., & Fried, E. I. (2018). A tutorial on regularized partial correlation networks. *Psychological Methods*, 23(4), 617–634. <https://doi.org/10.1037/met0000167>
- Epskamp, S., Maris, G., Waldorp, L. J., & Borsboom, D. (2018). Network psychometrics. In P. Irwing, T. Booth, & D. Hughes (Eds.), *The Wiley Handbook of Psychometric Testing: A Multidisciplinary Reference on Survey, Scale and Test Development* (pp. 953–986). John Wiley & Sons Ltd.
- Fritz, J., Fried, E. I., Goodyer, I. M., Wilkinson, P. O., & Harmelen, A.-L. V. (2018). A network model of resilience factors for adolescents with and without exposure to childhood adversity. *Scientific Reports*, 8(1), 1–13. <https://doi.org/10.1038/s41598-018-34130-2>
- Funkhouser, C. J., Correa, K. A., Gorka, S. M., Nelson, B. D., Phan, K. L., & Shankman, S. A. (2020). The replicability and generalizability of internalizing symptom networks across five samples. *Journal of Abnormal Psychology*, 129(2), 191–203. <https://doi.org/10.1037/abn0000496>
- Haslbeck, J. M. (2020). *Estimating group differences in network models using moderation analysis*. <https://psyarxiv.com/926pv>
- Haslbeck, J. M., Borsboom, D., & Waldorp, L. J. (2019). Moderated network models. *Multivariate Behavioral Research*, 56(2), 256–287. <https://doi.org/10.1080/00273171.2019.1677207>
- Hastie, T., Tibshirani, R., & Wainwright, M. (2015). *Statistical learning with sparsity: The lasso and generalizations*. CRC Press.
- Heeren, A., & McNally, R. J. (2018). Social anxiety disorder as a densely interconnected network of fear and avoidance for social situations. *Cognitive Therapy and Research*, 42(1), 103–113. <https://doi.org/10.1007/s10608-017-9876-3>
- Hoffman, M., Steinley, D., Trull, T. J., Lane, S. P., Wood, P. K., & Sher, K. J. (2019). The influence of sample selection on the structure of psychopathology symptom networks: An example with alcohol use disorder. *Journal of Abnormal Psychology*, 128(5), 473–486. <https://doi.org/10.1037/abn0000438>
- Lazarov, A., Suarez-Jimenez, B., Levi, O., Coppersmith, D. D. L., Lubin, G., Pine, D. S., Bar-Haim, Y., Abend, R., & Neria, Y. (2020). Symptom structure of PTSD and co-morbid depressive symptoms—a network

- analysis of combat veteran patients. *Psychological Medicine*, 50(13), 2154–2117. <https://doi.org/10.1017/S0033291719002034>
- Leroux, B. G. (1992). Consistent estimation of a mixing distribution. *The Annals of Statistics*, 20(3), 1350–1360. <https://doi.org/10.1214/aos/1176348772>
- Levinson, C. A., Brosof, L. C., Vanzhula, I., Christian, C., Jones, P., Rodebaugh, T. L., Langer, J. K., White, E. K., Warren, C., Weeks, J. W., Menatti, A., Lim, M. H., & Fernandez, K. C. (2018). Social anxiety and eating disorder comorbidity and underlying vulnerabilities: Using network analysis to conceptualize comorbidity. *International Journal of Eating Disorders*, 51(7), 693–709. <https://doi.org/10.1002/eat.122890>
- McLachlan, G. J., & Peel, D. (2000). *Finite mixture models*. Wiley.
- Meier, M., Kossakowski, J. J., Jones, P. J., Kay, B., Riemann, B. C., & McNally, R. J. (2019). Obsessive–compulsive symptoms in eating disorders: A network investigation. *International Journal of Eating Disorders*, 53(3), 362–371. <https://doi.org/10.1002/eat.23196>
- Mkhitarian, S., Crutzen, R., Steenaert, E., & Vries, N. K. D. (2019). Network approach in health behavior research: How can we explore new questions? *Health Psychology and Behavioral Medicine*, 7(1), 362–384. <https://doi.org/10.1080/21642850.2019.1682587>
- Mullarkey, M. C., Marchetti, I., Bluth, K., Carlson, C. L., Shumake, J., & Beevers, C. G. (2021). Symptom centrality and infrequency of endorsement identify adolescent depression symptoms more strongly associated with life satisfaction. *Journal of Affective Disorders*, 289, 90–97.
- Muthén, B. O. (1989). Latent variable modeling in heterogeneous populations. *Psychometrika*, 54(4), 557–585. <https://doi.org/10.1007/BF02296397>
- Pearl, J. (2009). *Causality* (2nd ed.). Cambridge University Press.
- Robinaugh, D. J., Hoekstra, R. H., Toner, E. R., & Borsboom, D. (2020). The network approach to psychopathology: A review of the literature 2008–2018 and an agenda for future research. *Psychological Medicine*, 50(3), 353–366. <https://doi.org/10.1017/S0033291719003404>
- Ryan, O., Bringmann, L. F., & Schuurman, N. K. (2019). The challenge of generating causal hypotheses using network models. *PsyArXiv*. <https://doi.org/10.31234/osf.io/ryg69>
- Saeed, B., Panigrahi, S., & Uhler, C. (2020). *Causal structure discovery from distributions arising from mixtures of DAGs*. arXiv. <https://arxiv.org/abs/2001.11940>
- Santos, H., Jr., Fried, E. I., Asafu-Adjei, J., & Ruiz, R. J. (2017). Network structure of perinatal depressive symptoms in Latinas: Relationship to stress and reproductive biomarkers. *Research in Nursing & Health*, 40(3), 218–228. <https://doi.org/10.1002/nur.21784>
- Schmittmann, V. D., Cramer, A. O., Waldorp, L. J., Epskamp, S., Kievit, R. A., & Borsboom, D. (2013). Deconstructing the construct: A network perspective on psychological phenomena. *New Ideas in Psychology*, 31(1), 43–53. <https://doi.org/10.1016/j.newideapsych.2011.02.007>
- Snoep, J. D., Morabia, A., Hernández-Díaz, S., Hernán, M. A., & Vandembroucke, J. P. (2014). Commentary: A structural approach to Berkson's fallacy and a guide to a history of opinions about it. *International Journal of Epidemiology*, 43(2), 515–521. <https://doi.org/10.1093/ije/dyu026>
- Southward, M. W., & Cheavens, J. S. (2018). Identifying core deficits in a dimensional model of borderline personality disorder features: A network analysis. *Clinical Psychological Science*, 6(5), 685–703. <https://doi.org/10.1177/2167702618769560>
- Spirtes, P., & Zhang, K. (2016). Causal discovery and inference: Concepts and recent methodological advances. *Applied Informatics*. (Vol. 3, No. 1, pp. 1–28). SpringerOpen.
- Steele, R. J., & Raftery, A. E. (2010). Performance of Bayesian model selection criteria for gaussian mixture models. *Frontiers of Statistical Decision Making and Bayesian Analysis*, 2, 113–130.
- van Rooijen, G., Isvoranu, A.-M., Kruijt, O. H., van Borkulo, C. D., Meijer, C. J., Wigman, J. T. W., Ruhé, H. G., & de Haan, L. (2018). A state-independent network of depressive, negative and positive symptoms in male patients with schizophrenia spectrum disorders. *Schizophrenia Research*, 193, 232–239. <https://doi.org/10.1016/j.schres.2017.07.035>

Received May 29, 2020

Revision received May 3, 2021

Accepted May 19, 2021 ■