Assessing the Collective Utility of Multiple Analyses On Clinical Alcohol Use Disorder Data

Erich Kummerfeld, PhD[1], Alexander Rix, MS[1],
Justin J. Anker, PhD[2], Matt G. Kushner, PhD[2]

[1]University of Minnesota – Institute for Health Informatics, Minneapolis, Minnesota
[2]University of Minnesota – Department of Psychiatry, Minneapolis, Minnesota

May 24, 2019

Abstract

Objective. The objective of this study was to assess the potential of combining graph learning methods with latent variable estimation methods for mining clinically useful information from observational clinical data sets.
Materials and Methods. The data set contained self-reported measures of psychopathology symptoms from a clinical sample receiving treatment for alcohol use disorder. We used the traditional graph learning methods Graphical Least Absolute Shrinkage and Selection Operator, and Friedman's hillclimbing algorithm; traditional latent variable estimation method factor analysis; recently developed graph learning method Greedy Fast Causal Inference; and recently developed latent variable estimation method Find One Factor Clusters. Methods were assessed qualitatively by the content of their findings.
Results. Recently developed graphical methods identified potential latent variables (i.e., not represented in the model) influencing particular scores. Recently developed latent effect estimation methods identified plausible cross-score loadings that were not found with factor analysis. A graphical analysis of individual items identified a mistake in wording on one questionnaire, and provided further evidence that certain scores are not reflective of indirectly measured common causes.
Discussion and Conclusion. Our findings suggest that a combination of Greedy Fast Causal Inference and Find One Factor Clusters can enhance the evidence-based information yield from psychopathological constructs and questionnaires. Traditional methods provided some of the same information, but missed other important findings. These conclusions point the way toward more informative interrogations of existing and future datasets than are commonly employed at present.

BACKGROUND AND SIGNIFICANCE

According to the 2015 National Survey on Drug Use and Health, alcohol use disorder (AUD) affects over 15 million people in the US alone, and in 2010 it was estimated that alcohol misuse cost the United States $249.0 billion[1]. Approximately one third of that population also suffers from anxiety or depression (internalizing) disorders, and following treatment, patients who suffer from both AUD and internalizing disorders are twice as likely to relapse following treatment[25]. As in many psychopathology domains, the mechanisms that produce and maintain comorbidity between these disorders are not well understood, so there is a critical need for discoveries that inform the prevention and treatment of AUD.

Despite the potential impacts of various mental health problems on AUD risk and recovery, they remain poorly understood as constructs. Even the ontological status of mental disorder symptoms as stemming from a latent cause (analogous to an unidentified infection) has recently been called into question by an alternative view that emphasizes networked interactions between specific manifest symptoms[6]. Given these fundamental and practical unknowns, psychopathology is in dire need of methodological advancements that can aid investigators in determining whether mental disorders have a single targetable latent source (analogous to an infection) or consist of an interacting network of symptoms (or a combination of the two[7]). This knowledge has the potential to reveal new insights leading to paradigm-changing innovations in diagnosis and treatment.

This project provides the first implementation and assessment of a novel analytic methodology that uses latent variable estimation methods (common cause models) in conjunction with graph learning methods (symptom

interaction models) to analyze questionnaire-based clinical mental health data at both the score level and item level. In doing so we assessed how the information produced by these different methods could be redundant, conflicting, and/or supplementary. It is beyond the scope of this work to provide an exhaustive coverage of graphical methods or latent variable methods. For example, 29 algorithms are offered in the graphical interface for Tetrad version 6.5.4, some of which contain dozens of more precisely specified algorithm options. Even more algorithms can be found in various R packages, MATLAB files, and elsewhere. To maximize feasibility and impact, we focus on the comparative and collective utility of selected traditional techniques in common use and a limited number of promising recent developments that are, as yet, little used in the field.

The graphical LASSO (GLASSO)[8] and Friedmans hillclimbing algorithm (FHC)[9,10] were included as traditional graph learning methods, as they have received significant attention and support in the psychopathology literature. Factor analysis (FA)[11] was included as the traditional and dominant latent variable estimation method of the last century. Greedy Fast Causal Inference (GFCI)[12] was included as a recently developed graph learning method due to its ability to determine the possible influence of latent common causes not represented in the model and its performance in simulation experiments at smaller sample sizes. Find One Factor Clusters (FOFC)[13] was included as a recently developed latent variable estimation method due to its speed and performance in simulation experiments at smaller sample sizes.

These methods do not rely strictly on Bayesian or frequentist conceptions of probability. While the term Bayesian network is often used to refer to probabilistic directed graphical models, such models can be evaluated in either a Bayesian or frequentist manner. Of the network-leaning algorithms covered in this paper, only FHC learns standard Bayesian networks: GLASSO learns an undirected graph and GFCI learns a more complex type of network.

MATERIALS AND METHODS

Data

Data were collected from 362 adult AUD treatment inpatients with a co-occurring anxiety disorder at the baseline assessment of a randomized clinical trial (RCT) [14]. The baseline assessment occurred prior to the RCT clinical interventions within the first week of a 21-day residential chemical dependency treatment program. Variables represent hallmarks of comorbidity between AUD and anxiety disorder and include measures of anxiety and depression (internalizing) symptoms, stress and coping abilities, drinking behaviors, and alcohol craving. Measures were self-reported and obtained from empirically validated assessments, with no skip questions, and the data set contained very few missing values, leading to a high quality data set. Variables were constructed from individual items based on standard scales for the various internalizing disorders.

Measure	Mean (SD)	Range	Description
Generalized anxiety	64.13 (11.59)	16-80	The total score on the Penn State Worry Questionnaire.
Depression	20.43 (17.30)	0-63	The total score on the Beck Depression Inventor.
Social anxiety	32.43 (17.30)	0-80	The total score on the Social Phobia Scale.
Panic	10.99 (6.34)	0-28	The total score on the Panic Disorder Severity Scale.
Agoraphobia	31.59 (19.78)	0-100	The summed score from the Mobility Inventory for Agoraphobia.
Perceived stress	28.15 (5.50)	10-40	The total score on the Perceived Stress Scale.
Self-efficacy	32.91 (10.91)	8-48	The total score on the negative affect subscale of the Situational Confidence Questionnaire.
Drinking to cope	62.93 (12.15)	20-80	The Unpleasant Emotions subscale of the Inventory of Drinking Situations.
Drinking behavior	1608.76 (1271.51)	30-6840	The total drinks consumed during the 4 months prior to residential treatment entry assessed with the Timeline Follow-Back Interview.
Alcohol craving	2.67 (1.05)	0 to 4	The frequency of alcohol craving during the 30 days prior to treatment assessed with an item from the Obsessive Compulsive Drinking Scale.

Table 1: Measures collected from AUD patients

Methods

Term	Description
AUD	Alcohol Use Disorder is a mental disorder characterized by excessive consumption of alcohol.
Causal Graph	A causal graph is a structural representation of the causal relationships among a set of variables.
Causal Relationship	Two variables A and B have a causal relationship if A causes B, or B causes A, or A and B are both caused by some third variable C.
DAG	A Directed Acyclic Graph is a graph whose nodes are only connected by arrows and that contains no directed cycles.
DTC	Drinking to Cope is a measure of drinking for coping purposes.
FA	Factor Analysis is a collection of methods for learning factor models from data.
FHC	Friedmans hillclimbing algorithm is a popular method for learning DAGs from data.
FOFC	Find One Factor Clusters is a recently developed algorithm for identifying latent common causes from data.
GFCI	Greedy Fast Causal Inference is a recently developed algorithm for learning PAGs from data.
GLASSO	Graphical Least Absolute Shrinkage and Selection Operator is a popular tool for learning undirected graphs from data.
GLEE	Graphical latent effect estimation is a collection of methods for identifying latent common causes from data.
Latent Common Cause	L is a latent common cause of variables A and B if L is latent variable and is a cause of both A and B.
Latent Variables	L is a latent variable in the context of data set D if L is not explicitly represented as a variable measured in D.
PAG	A Partial Ancestral Graph contains many more types of edges than a DAG (see Table 3), and is able to represent certain or possible confounding by latent variables. PAGs also do not contain cycles.

Table 2: Glossary of terms and abbreviations

Graph Learning Methods

GLASSO[8] estimates a regularized inverse covariance matrix, which can be represented as an undirected graph. It is a popular tool for discovering unoriented graphs from observational data, including data related to psychopathology[25,26]. The primary distinction between GLASSO and the other graph-learning algorithms we employed is that GLASSO does not encode causal information, so it serves as a point of comparison for the causal methods. Undirected graphs can be difficult to interpret, especially as the number of variables and edges increases. Because of this, their interpretation is typically done at a relatively high level: the graph is fed into an analysis method which evaluates various graphical metrics, such as the centrality and connectedness of each node. The nodes that rank highly on these metrics are identified as being important to maintaining connectivity in the network. Well-connected groups of nodes (clusters) can also be identified as collections of variables that seem to be categorically similar. We utilized the version of GLASSO implemented in the R[27] package glasso.

FHC[9,10] has recently been utilized in some psychopathology publications[28]. This method models the causal structure of the variables by optimizing a complexity-penalized likelihood score, typically the Bayesian Information Criterion (BIC). It has not been proven to be correct in the infinite data limit, but can perform well in simulations. It outputs a directed acyclic graph (DAG) where nodes are connected by arrows. The directed edges in these graphs are frequently interpreted causally, such that the edge A B is interpreted to mean that the variable A causally influences the variable B, as utilized by Pearl[29] and Spirtes[30], and explicated by Woodward[31]. As a consequence of outputting a DAG, directionality is forced, and the possible influence of latent variables is not accounted for. There are numerous other methods that produce directed, or partially directed, graphical models such as PC[30], GES[32], and MMHC[33]; we selected FHC due to its use in prior psychopathology research. We utilized the version of FHC implemented in the R package bnlearn[10].

GFCI[12] uses a combination of penalized likelihood score comparisons and conditional independence tests to learn the relationships among variables while accounting for possible latent common causes. It has been proven

correct in the infinite data limit, and while benchmark simulations of its performance on finite sample sizes are as-yet limited in scope, the benchmarking that has been done so far is promising[12]. In terms of scalability to large numbers of variables, more complex algorithms are naturally slower than methods like GLASSO, however there are many data sets, such as the one covered in this paper, which are well within GFCI's feasibility bounds. GFCI outputs a partial ancestral graph (PAG)[30], a graphical representation that encodes the possibility of latent common causes. PAGs use a rich set of edge types to encode a large amount of information, including whether a given relationship is definitely, possibly, or definitely not confounded, as well as whether a variable definitely, possibly, or definitely does not cause another variable. In the typical representation, the inclusion of a circle at either end of the edge (e.g. o-o or o-¿) indicates the possibility that a latent common cause may be responsible for part (or all) of the statistical signal between those variables, while the inclusion of arrowheads at both ends of the edge indicates that the relationship is definitely due to a latent common cause. Table 3 is a reference for the meaning of edges in a PAG. We used the only current implementation of GFCI, found in the Tetrad java software package.

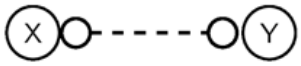
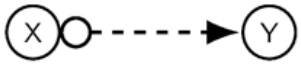
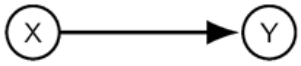
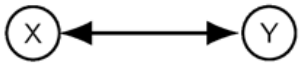
Edge Type	Meaning
	Precisely one of the following is true: a. X causes Y b. Y causes X c. X and Y are confounded d. both a and c e. both b and c
	Y is not a cause of X. In addition, at least one of the following is true: a. X causes Y b. X and Y are confounded
	All of the following are true: a. X is a direct or indirect cause of Y. b. X and Y are not confounded. c. Y is not a cause of X.
	All of the following are true: a. X is a direct cause of Y. b. X and Y are not confounded. c. Y is not a cause of X.
	All of the following are true: a. There is a latent common cause of X and Y. b. X is not a direct cause of Y. c. Y is not a direct cause of X.

Table 3: Edge types in a Partial Ancestral Graph (PAG)

Latent Variable Estimation Methods

There are two primary approaches for estimating latent common causes: factor analysis (FA)[11] and graphical latent effect estimation (GLEE)[13,34,35]. FA is the traditional and dominant statistical method used for this purpose over the last century, and it continues to be widely used. It attempts to estimate the data as a product of two smaller matrices, and it makes inferences based on the structural features of those matrices. The latents identified by factor analysis do not necessarily have a causal interpretation, in the sense that factor models remain ambiguous as to whether factors cause the measured variables or are defined with them, but rather serve as a simplified description of the covariances observed in the data. We utilized the version of FA implemented in the factanal function in R.

GLEE methods have only been developed in the last 15 years[13,35], and remain largely untested. Unlike

factor analysis, GLEE methods explicitly aim to identify the existence of latent variables that causally influence the measured variables. We used a particular GLEE method, the Find One Factor Clusters (FOFC)[13] algorithm, because of its simulation performance and familiarity to the authors. FOFC uses vanishing tetrad tests to determine which, if any, subsets of the measured variables are correlated by a single latent common cause. While FOFC has been proven correct and performs well in simulations at sample sizes comparable to that of our data set, its output on finite sample sizes can depend on the variable order, so we ran FOFC 100 times on random variable orderings and stored all of the output factors that it identified. Output factors that loaded onto a small number of items (¡ 5) were dropped since small factors are less reliable than larger factors for this method[13]. K-means clustering (with k=4 selected to mirror the choice of 4 factors) was used to aggregate the output from the 100 FOFC runs. We utilized the only current implementation of FOFC, found in the Tetrad java software package. Table 4 summarizes the similarities and differences across all methods used in this paper.

Learning Method	Representation	Causal Intrepretation	Latents	Correctness Proof
GLASSO	Undirected graph	no	no	yes
Hillclimbing	DAG	yes	no	no
GFCI	PAG	yes	allowed	yes
Factor analysis	Factor model	no	modeled	no
FOFC	Latent variable model	yes	modeled	yes

Table 4: Comparison of utilized learning methods

RESULTS

Application to Clinical Data

We first consider the undirected graph learned by GLASSO[36], shown in Figure 1. The graph shows dense connections among the various internalizing disorders, and also has the strongest connections from the alcohol focused variables (Drinking and Craving) to the Drinking to Cope (DTC) variable. Our analysis replicated earlier results of applying GLASSO to this data set[36], including DTC receiving the highest graphical metrics of betweenness, closeness, and strength. These heuristics indicate that DTC could play an important role in the common co-occurrence of drinking problems and internalizing disorders, but they do not contain causal information.

To begin investigating the causal relationships among the variables in this data set we used Friedmans hill-climbing algorithm (FHC)[10] to learn a directed acyclic graph (DAG), shown in Figure 2. The DAG encodes a variety of descriptive statistics and causal information that is absent from the GLASSO model. For example, according to this DAG, conditioning or controlling for DTC makes Drinking statistically independent of Depression. For causal information, the DAG implies that effectively treating someone's depression would not affect their drinking, but effectively treating their social anxiety, stress, and improving their self-efficacy to resist DTC would.

To investigate the possibility of latent common causes (confounders), we applied Greedy Fast Causal Inference (GFCI)[12], which produced the partial ancestral graph (PAG) shown in Figure 3. We validated this PAG, and all other PAGs in this paper, using a jackknife procedure (see Online Supplement Tables S3, S4, S5, and S6). Table 3 provides the interpretation of the edges found in this PAG. This graph confirms some of the information contained in the DAG, such as the causal chain DTC Craving Drinking, but disagrees with the DAG in other areas, such as whether DTC causes self-efficacy or self-efficacy causes DTC. All of the variables that are connected to each other in the PAG are also connected in the DAG, but the DAG contains a small number of adjacencies that are not present in the PAG. These differences could be errors induced by latent common causes, but could also be due to differences in the parameter settings we used for FHC and GFCI: the two methods have distinct sets of parameters with different interpretations and meanings, so it is difficult to determine if they are equivalently calibrated.

There are a few striking similarities between the DAG and the PAG. DTC is causally upstream of Craving and Drinking in both graphs, and GFCI goes a step further by confirming that these relationships are not confounded by latent variables. GFCI also discovers that the causal effect of Social Anxiety on DTC is not confounded, but the causal relationship between Stress and DTC might be: in fact it allows for the possibility that Stress is not a cause of DTC at all, but rather their correlation could be due entirely to a latent common cause.

GFCI identified that half of the causal relationships in the PAG are possibly confounded. This casts doubt on the orientations that FHC gave to those edges. The possibility of confounding is especially present within the distress domain (Depression and General Anxiety), as all of the causal relationships connecting these variables to other variables in the graph are possibly confounded. There is also a small number of orientation disagreements between the GFCI PAG and the DAG learned by FHC, but these are simply low confidence graph features.

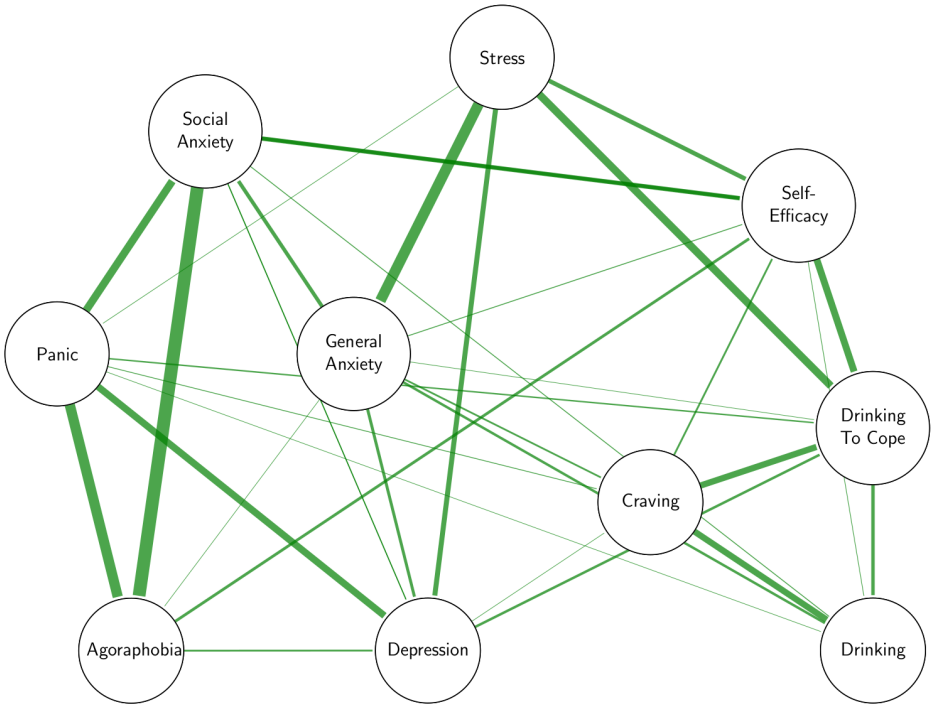


Figure 1: Visualization of graph learned with GLASSO.

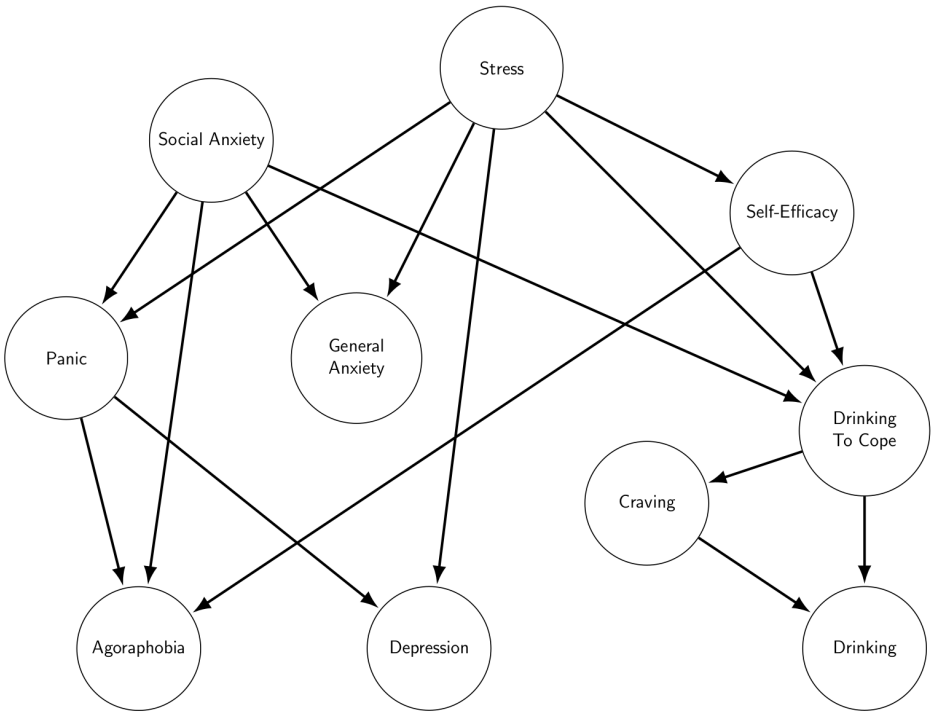


Figure 2: Visualization of graph learned with FHC.

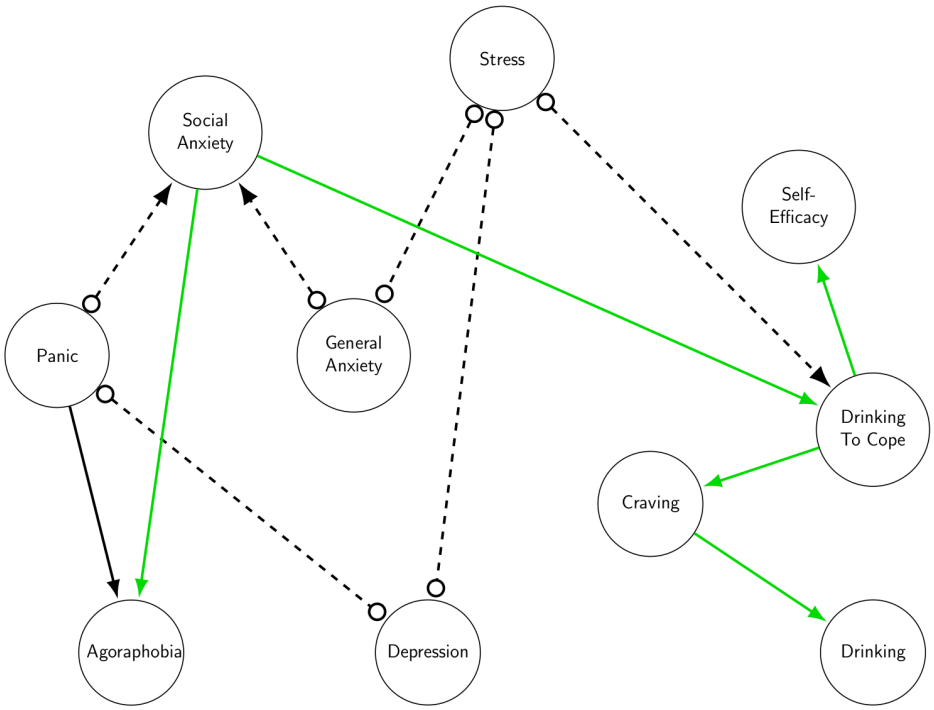


Figure 3: PAG of the measured scores described in Table 1. The different types of edges, such as green arrows, are described in Table 2. Latent variables are not represented explicitly, but their potential influence is captured in the choice of edge type connecting 2 nodes.

Our psychopathologists were surprised at the prominence of Social anxiety within the causal networks over other possible causes of Drinking. Both models indicated that Social Anxiety served as an initiator of the causal chain of conditions that terminate with Drinking, and GFCI confirmed that this entire causal pathway was unconfounded.

Factor Analysis of Item-Level Data

After GFCI identified several pairs of score-level variables which might be confounded, we hypothesized that latent common causes of the total scores were operating on them indirectly through their items. To test this hypothesis, we analyzed the item-level data for the Stress, General Anxiety, Depression, and Panic scores, as the PAG indicated that all of the relationships among these scores might be confounded. This created a data set of 54 items: 10 for Stress, 21 for Depression, 7 for Panic, and 16 for General Anxiety.

We applied factor analysis to the item-level data, using four factors since we know they come from four scores/assessments. This was augmented with the oblique promax rotation, since the scores are correlated (as shown by the undirected graph, the DAG, and the PAG). The results are shown in Online Supplement Table S1. The factor analysis does not identify any cross-loadings, even at a low cutoff threshold of 0.3, which would imply that the items are all measuring only the factors/scores that they are intended to.

For a second perspective, we also applied FOFC[13] to the same item-level data set. The cluster centroids are shown in Online Supplement Table S2. In contrast to the factor analysis, FOFC finds that some cross-score items share a latent common cause. In particular, centroid 4 contains one General Anxiety item sharing a latent cause with 4 Stress items. Our psychopathology experts inspected these test items and were excited to find that they were related to feelings of being overwhelmed by the daily obligations and routines of life, content which is largely missing from the other General Anxiety and Stress items and is very relevant to the lives of patients with AUD.

Graphical Analysis of Item-level Data

To further investigate the patterns in our data, we performed graphical analyses on targeted sets of questionnaire items. These sets of questionnaire items are typically summed into a score, as in our earlier analyses. Doing so implicitly treats the items as being reflective of a latent common cause or forming a construct of interest. This sort of analysis has recently been challenged, however [6], and the results of the FOFC analysis suggest that at a minimum, some of these scores do not appear to correspond to a latent common cause. Since FOFC rejected

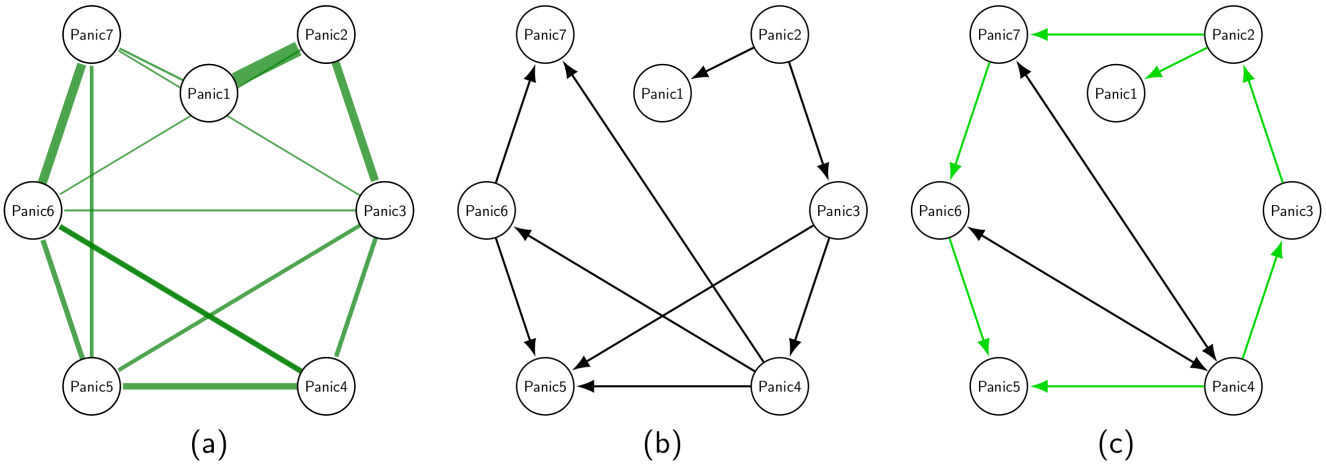


Figure 4: Graphical representations of the item structure of the Panic Disorder Severity Scale. Table 2 provides descriptions of the edges in representation (c). Each item assesses a different aspect of panic experienced during the past 30 days and include the following: (1) number of panic attacks, (2) severity of panic attacks, (3) anxiety about next panic attack, (4) avoiding places due to fear of panic attacks, (5) avoiding activities due to fear of or similarity to panic attacks, (6) the extent panic-related symptoms interfere with work and other responsibilities, (7) the extent panic-related symptoms interfere with social life.

the possibility that the Panic items formed an ideal latent measurement model, we investigated the empirical relationships among the Panic items with GLASSO, FHC, and GFCI. The results are shown in Figure 4.

Figure 4 (a) shows the results of GLASSO applied to the Panic items, which indicates that some of the items are much more strongly connected than others. Figure 4 (b) shows the result of the FHC analysis. FHC produces a more sparse graph than GLASSO, and structurally appears similar to the GLASSO graph after removing the weakest edges. FHC expresses a number of direct causal relationships, but since it cannot represent latent variables or their statistical influence on measured variables, this is not itself evidence of the absence of a latent common cause.

The GFCI analysis output is shown in Figure 4 (c). It contains a mixture of unconfounded direct causal relationships and latent common cause relationships. The numerous unconfounded direct causal relationships explains the empty section of the FOFC output table: FOFC does not group variables together into a measurement model if they have direct causal relationships. These questionnaire items ask the respondent to provide information about the frequency, severity, and functional consequences of the respondent's panic attacks, and are written in a way that assumes the respondent is experiencing panic on a regular basis. It seems plausible that some of these items would directly influence other items. The bi-directed edges between items Panic6 and Panic4, and items Panic7 and Panic4, also make sense: Panic4 asks if the respondents panic attacks hinder their ability to travel, and Panic6 and Panic7 ask if the respondents panic attacks hinder their ability to go to work or socialize, both being activities that often involve traveling.

Unlike Panic, large portions of the Depression items were grouped together by both FA and FOFC. Reviewing the questionnaire revealed that unlike with Panic, these items do seem designed as indicators rather than queries about the breadth and severity of symptoms of an assumed disease. Combined with the factor analysis and FOFC analysis, this suggests that these items may form a reflective model of underlying Depression, and so we applied our graphical techniques to these items to serve as a counterpoint to the Panic items. Figure 5 shows the results of applying (a) GLASSO, (b) FHC, and (c) GFCI to the Depression questionnaire items.

The most notable difference between the GLASSO, FHC, and GFCI analyses of the Depression items is that the GLASSO and GFCI graphs are dense, having a large number of edges and being difficult to untangle visually, while the FHC graph is sparse, with very few edges and several nodes having no connections at all. Exactly why FHCs output differs so substantially from the other algorithms in this particular case is unclear. One possibility is that it relates to the parameters we are using when running the FHC algorithm, however these are not different than what we have used for every other FHC analysis presented in this paper.

The GFCI graph also highlights Dep11 as an item deserving additional investigation. These questionnaires are designed in such a way that the items are supposed to be correlated, and so by design there should be no unconnected nodes in an accurate graphical representation of the items, however Dep11 is unconnected in the GFCI graph. Upon inspecting the questionnaire, we identified an error in the final, printed version of the Dep11 question, which resulted in it being uncorrelated with the other items. This demonstrates an interesting application of GFCI and likely other advanced graphical methods for verifying or identifying errors in psychopathology questionnaires.

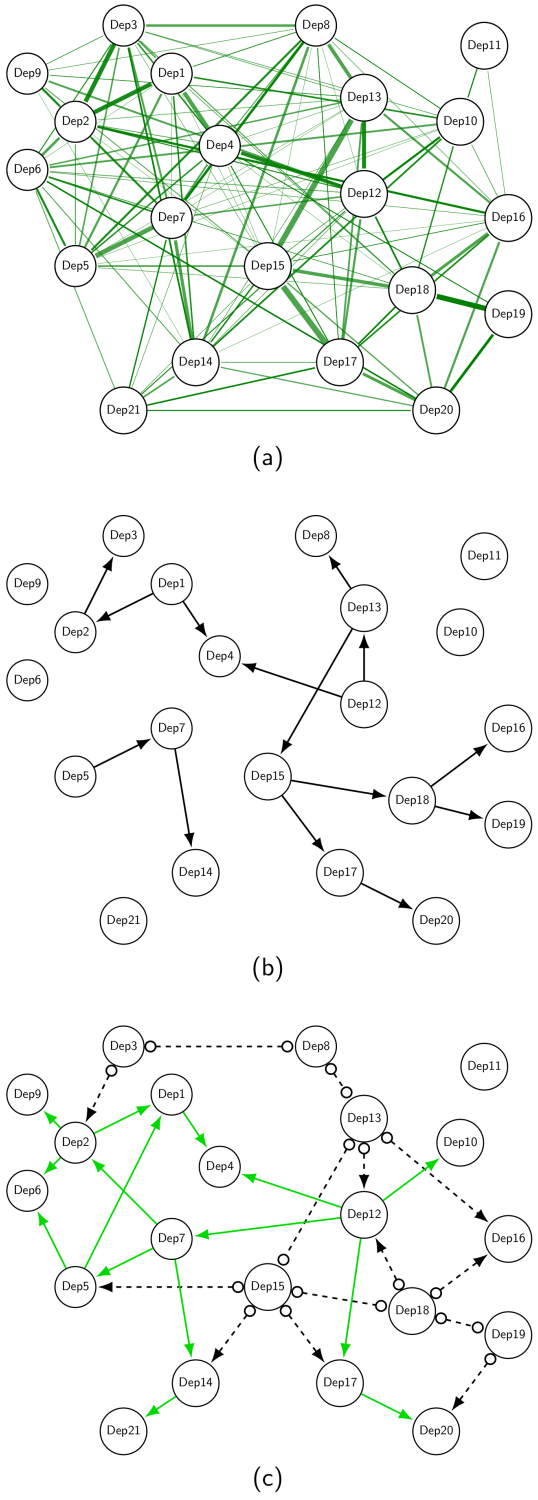


Figure 5: Graphical representations of the item structure of the Beck Depression Inventory. Table 2 provides descriptions of the edges in representation (c). Items assess the following aspects of depression: (1) feelings of sadness, (2) hopelessness, (3) feeling like a failure, (4) feeling dissatisfied, (5) feeling guilty, (6) feeling like they are being punished, (7) self-hate, (8) self-blame, (9) suicidal intent, (10) amount of crying, (11) feelings of irritation, (12) lost interest in other people, (13) inability to make decisions, (14) feeling unattractive, (15) inability to work, (16) difficulty sleeping, (17) tiredness, (18) lack of appetite, (19) weight loss, (20) worry about physical health, (21) lost interest in sex.

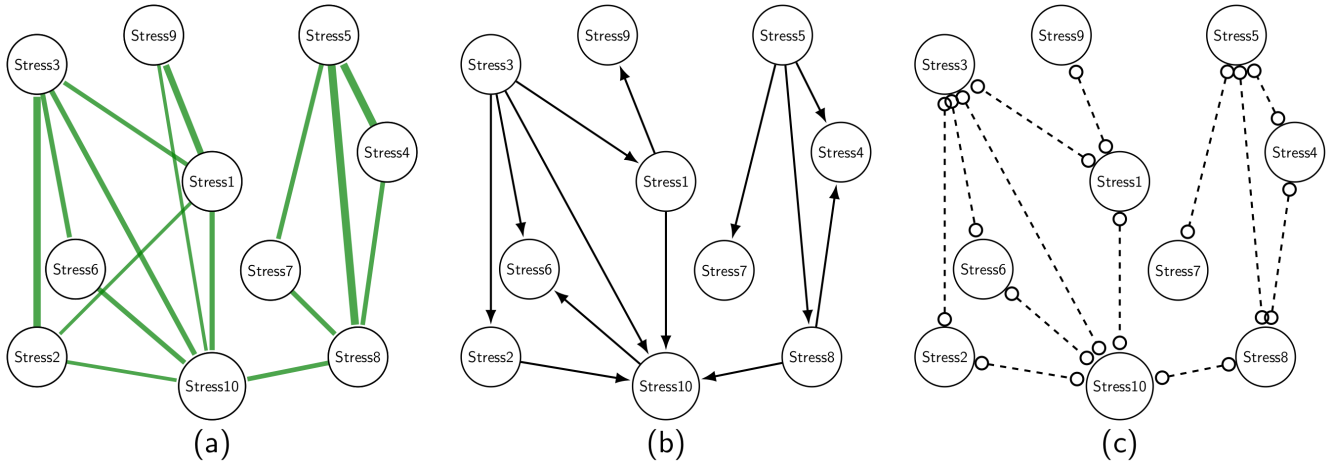


Figure 6: Graphical representations of the item structure of the Perceived Stress Scale. Table 2 provides descriptions of the edges in representation (c). Each item measures a different aspect of stress that include the following: (1) upset by unexpected events, (2) feeling unable to control important things, (3) feeling nervous and stressed, (4) felt confident about handling personal problems (reversed), (5) felt things were going your way (reversed), (6) could not cope with responsibilities, (7) able to control irritations in life (reversed), (8) felt that you were on top of things (reversed), (9) angered by things outside your control, (10) felt unable to overcome difficulties.

Notably, such errors did not exist for the other nodes which are unconnected in the FHC graph only.

We also investigated the empirical patterns among the Stress questionnaire items, since (1) the score-level analysis identified Stress as an important variable, and (2) the factor analysis and FOFC analysis produced substantially different results on the stress questionnaire items. This collectively suggests stress is both important and perhaps poorly measured. Visualizations of the Stress item analyses performed by GLASSO, FHC, and GFCI are shown in Figure 6.

Figure 6 (a) shows the GLASSO analysis of the Stress items. It indicates that the variables may be loosely divided into 2 groups, with overall stronger edge strength between variables within each group and weaker edge strength for edges spanning the two different groups. The output of FHC, shown in Figure 6 (b), also divides the variables roughly into two groups. As usual, every edge in this graph is oriented, but since we have reason to believe that there may be latent common causes of the questionnaire items, these edge orientations cannot be safely interpreted as causal relationships.

GFCIs analysis is shown in Figure 6 (c), and again shows the items roughly divided into two groups. Unlike the Panic questionnaire items, GFCI does not identify any unconfounded direct causal relationships. The presence of nothing but double-circle edges is consistent with the presence of one or more latent common causes influencing the items, although it also does not serve as evidence of the presence of latent common causes. It does, however, imply that the edge orientations of the FHC analysis should not be trusted: they could very well be the product of FHCs representational limitations.

It is noteworthy that the division of variables into roughly two groups is consistent across all three analytical methods. Upon inspection of the items, we identified that the two groups correspond precisely to whether an item is reverse-scored or not. Reverse-scored items are those where higher values indicate lower levels of stress, as compared to the regular items where higher values indicate higher levels of stress. In the data, the reverse-scored items had their values reversed, so that (1) the higher values correspond to increased stress, and (2) the reversed and non-reversed items are expected to be positively correlated with each other.

The graphical separation of the items into these two groups suggests that whether an item was reversed or not is important to the way respondents answer that item. In particular, the correlations among the items may be strongly influenced by psychological effects of trying to answer a reversed or non-reversed question. Regarding whether this Stress score reflects a common cause or is the product of a symptom network, there is mixed evidence. On the one hand, no unconfounded edges are identified, which is indicative of a possible latent common cause. On the other hand, the graph is clearly divided between the reversed and non-reversed items, suggesting that all of these items do not share a single latent common cause. Alternative theories that should be pursued are: (a) there are two latent common causes, which are somehow being separately targeted by the reversed and non-reversed items; (b) participants may be psychologically responding more to how the questions are framed—i.e. whether the questions are stated in a reversed way or not—than to the content of the questions themselves.

DISCUSSION

Although Galton pioneered the use of questionnaires to investigate psychological phenomena in the 19th century, and Spearman developed factor analysis to explain correlated test performances in the early 20th century, most of the methods used in this paper were developed in the 21st century. The more recent methods still make use of 19th century concepts such as correlation and regression, but they have either significantly refined them (GLASSO) or incorporated intelligent search procedures to guide a long sequence of different statistical procedures (FHC, GFCI, FOFC).

In addition to replicating known findings, (e.g. that DTC causes Craving, and that Craving causes Drinking[37]) new knowledge was produced by the recently developed methods applied in this work. For example, we found new evidence that Social Anxiety is partly responsible for the co-occurrence of depression and AUD and that DTC, Stress, and Social Anxiety may be the key treatment targets that would impact down-stream drinking behavior.

The Social Anxiety DTC Craving Drinking causal chain is supported by findings from several other studies. In one study, Social Anxiety was highly correlated with endorsement of drinking in unpleasant emotions and, in fact, DTC mediated the relationship between Social Anxiety and problematic alcohol consumption[38]. In another study of a community sample, DTC mediated the relationship between Social Anxiety and drinking problems in those with AUD[39]. In still another study of individuals with AUD, drinking to cope mediated the relationship between Social Anxiety and drinking problems[40].

Although the primary purpose of this paper was to investigate the information provided by multiple methodologies applied to the same dataset, it contains a number of models related to real-world clinical phenomena. We caution readers to limit their interpretation of these models, which may contain errors or fail to generalize. The reasons for this include a relatively small sample size, untuned learning parameters, treating ordinal variables as linear variables, the possibility of causal cycles, and being restricted to subjects entering a treatment program. For published models learned from clinical psychopathology data such limitations are common.

Our findings demonstrate that recently developed analytic methods can extract useful knowledge from data sets that traditional analyses do not. Unlike GLASSO and FHC, GFCI identified that the causal chain from social anxiety to drinking is not confounded by the influence of latent variables. As an unexpected but revealing finding, GFCI uniquely brought a wording error that changed the meaning of a Depression item to our attention. FOFC revealed that despite the impressive factor analysis results, some of the items may have unwanted direct or indirect statistical relationships. The Panic questionnaire in particular was revealed to be problematic. More work must be done to develop improved questionnaire items for measuring these psychopathological constructs. We also hope that future algorithmic developments overcome some remaining obstacles; e.g. GFCI is unable to model causal cycles without time information, and FOFC is unable to identify useful but imperfect measurement models.

CONCLUSION

In this paper we demonstrated that a combination of traditional and recently developed graph learning methods and latent variable estimation methods applied to a clinical data set of patients with alcohol use disorder enables the discovery of new and important knowledge. Such knowledge included not only clinical insights about the treatment of patients with alcohol use disorder, but also previously unknown impactful errors in the tests being used. The present work could be directly extended by analyzing other data sets that use the same items, or potentially by constructing new sets of items intended specifically to target these latent variables. Moreover, these methods could easily and beneficially be applied to other clinical domains, as well as to more biological domains such as protein signalling[41].

Acknowledgements

The authors acknowledge the Center for Causal Discovery for the development of the statistical software (Tetrad) used by Erich Kummerfeld and Alexander Rix to conduct the GFCI and FOFC analyses. The authors thank Constantin Aliferis for helpful discussions and feedback on an earlier version of this manuscript. The authors also thank an anonymous associate editor and an anonymous reviewer for their helpful criticism and feedback.

Funding Statement

This work was supported by NIAAA grant number K01AA024805, NIAAA grant number R01-AA015069, NIH/NIMH grant number 1R01MH116156-01A1, and NIH/NCATS grant number UL1TR002494.

Competing Interests Statement

The authors have no competing interests to declare.

Contributorship Statement

All authors made substantial contributions to the conception and design of the work presented in this paper. MGK and JAA acquired and interpreted the data, and EK and AR ran the analysis. The paper was drafted by EK, and all authors revised it critically for important intellectual content, provided final approval of the version to be published, and agree to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

REFERENCES

1. Sacks JJ, Gonzales KR, Bouchery EE, Tomedi LE, Brewer RD. 2010 National and State Costs of Excessive Alcohol Consumption. Am J Prev Med. 2015;49: e73e79.
2. Greenfield SF, Weiss RD, Muenz LR, Vagge LM, Kelly JF, Bello LR, et al. The effect of depression on return to drinking: a prospective study. Arch Gen Psychiatry. 1998;55: 259265.
3. Kaufmann CN, Chen L-Y, Crum RM, Mojtabai R. Treatment seeking and barriers to treatment for alcohol use in persons with alcohol use disorders and comorbid mood or anxiety disorders. Soc Psychiatry Psychiatr Epidemiol. 2014;49: 14891499.
4. Kushner MG, Abrams K, Thuras P, Hanson KL, Brekke M, Sletten S. Follow-up study of anxiety disorder and alcohol dependence in comorbid alcoholism treatment patients. Alcohol Clin Exp Res. 2005;29: 14321443.
5. Schellekens AFA, de Jong CAJ, Buitelaar JK, Verkes RJ. Co-morbid anxiety disorders predict early relapse after inpatient alcohol treatment. Eur Psychiatry. 2015;30: 128136.
6. Borsboom D, Cramer A, Kalis A. Brain disorders? Not really Why network structures block reductionism in psychopathology research. Behav Brain Sci. 2018; 154.
7. Fried EI, Cramer AOJ. Moving Forward: Challenges and Directions for Psychopathological Network Theory and Methodology. Perspect Psychol Sci. 2017;12: 9991020.
8. Friedman J, Hastie T, Tibshirani R. Sparse inverse covariance estimation with the graphical lasso. Biostatistics. 2008;9: 432441.
9. Friedman N, Goldszmidt M, Wyner A. Data Analysis with Bayesian Networks: A Bootstrap Approach. Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.; 1999. pp. 196205.
10. Scutari M. Learning Bayesian Networks with the bnlearn R Package. Journal of Statistical Software, Articles. 2010;35: 122.
11. Krueger RF. The structure of common mental disorders. Arch Gen Psychiatry. 1999;56: 921926.
12. Ogarrio JM, Spirtes P, Ramsey J. A Hybrid Causal Search Algorithm for Latent Variable Models. JMLR Workshop Conf Proc. 2016;52: 368379.
13. Kummerfeld E, Ramsey J. Causal Clustering for 1-Factor Measurement Models. KDD. 2016;2016: 16551664.
14. Kushner MG, Maurer EW, Thuras P, Donahue C, Frye B, Menary KR, et al. Hybrid cognitive behavioral therapy versus relaxation training for co-occurring anxiety and alcohol disorder: a randomized clinical trial. J Consult Clin Psychol. 2013;81: 429442.
15. Meyer TJ, Miller ML, Metzger RL, Borkovec TD. Development and validation of the Penn State Worry Questionnaire. Behav Res Ther. 1990;28: 487495.
16. Beck AT, Steer RA, Brown GK. BDI-II: Beck depression inventory. Pearson; 1996.
17. Mattick RP, Clarke JC. Development and validation of measures of social phobia scrutiny fear and social interaction anxiety. Behav Res Ther. Elsevier; 1998;36: 455470.
18. Houck PR, Spiegel DA, Shear MK, Rucci P. Reliability of the self-report version of the panic disorder severity scale. Depress Anxiety. 2002;15: 183185.
19. Chambless DL, Caputo GC, Jasin SE, Gracely EJ, Williams C. The Mobility Inventory for Agoraphobia. Behav Res Ther. 1985;23: 3544.
20. Cohen S, Kamarck T, Mermelstein R. A global measure of perceived stress. J Health Soc Behav. 1983;24: 385396.
21. Annis H, Martin Graham J, Addiction Research Foundation, Foundation OAR. Situational Confidence Questionnaire (SCQ): Users Guide. Addiction Research Foundation; 1988.
22. Annis HM, Graham JM. Profile types on the Inventory of Drinking Situations: Implications for relapse prevention counseling. Psychol Addict Behav. Educational Publishing Foundation; 1995;9: 176.
23. Sobell LC, Sobell MB. Timeline Follow-Back. Measuring Alcohol Consumption. Humana Press, Totowa, NJ; 1992. pp. 4172.
24. Anton RF, Moak DH, Latham PK. The obsessive compulsive drinking scale: A new method of assessing outcome in alcoholism treatment studies. Arch Gen Psychiatry. 1996;53: 225231.
25. McNally RJ. Can network analysis transform psychopathology? Behav Res Ther. 2016;86: 95104.
26. Epskamp S, Borsboom D, Fried EI. Estimating psychological networks and their accuracy: A tutorial paper. Behav Res Methods. 2018;50: 195212.
27. Team RC. R: A language and environment for statistical computing.

28. McNally RJ, Mair P, Mugno BL, Riemann BC. Co-morbid obsessive-compulsive disorder and depression: a Bayesian network approach. Psychol Med. 2017;47: 12041214.
29. Pearl J. Causality. Cambridge University Press; 2009.
30. Spirtes P, Glymour CN, Scheines R, Heckerman D, Meek C, Cooper G, et al. Causation, Prediction, and Search. MIT Press; 2000.
31. Woodward J. Making Things Happen: A Theory of Causal Explanation. Oxford University Press, USA; 2005.
32. Chickering DM. Optimal Structure Identification With Greedy Search. J Mach Learn Res. 2002;3: 507554.
33. Tsamardinos I, Brown LE, Aliferis CF. The max-min hill-climbing Bayesian network structure learning algorithm. Mach Learn. 2006;65: 3178.
34. Kummerfeld E, Ramsey J, Yang R, Spirtes P, Scheines R. Causal Clustering for 2-Factor Measurement Models. Machine Learning and Knowledge Discovery in Databases. Springer Berlin Heidelberg; 2014. pp. 3449.
35. Silva R, Scheine R, Glymour C, Spirtes P. Learning the Structure of Linear Latent Variable Models. J Mach Learn Res. 2006;7: 191246.
36. Anker JJ, Forbes MK, Almquist ZW, Menk JS, Thuras P, Unruh AS, et al. A network approach to modeling comorbid internalizing and alcohol use disorders. J Abnorm Psychol. 2017;126: 325339.
37. Anker JJ, Kummerfeld E, Rix A, Burwell SJ, Kushner MG. Causal Network Modeling of the Determinants of Drinking Behavior in Comorbid Alcohol Use and Anxiety Disorder. Alcohol Clin Exp Res. 2018; doi:10.1111/acer.13914
38. Buckner JD, Eggleston AM, Schmidt NB. Social anxiety and problematic alcohol consumption: the mediating role of drinking motives and situations. Behav Ther. 2006;37: 381391.
39. Clerkin EM, Sarfan LD, Parsons EM, Magee JC. Mindfulness Facets, Social Anxiety, and Drinking to Cope with Social Anxiety: Testing Mediators of Drinking Problems. Mindfulness. 2017;8: 159170.
40. Cooper R, Hildebrandt S, Gerlach AL. Drinking motives in alcohol use disorder patients with and without social anxiety disorder. Anxiety Stress Coping. 2014;27: 113122.
41. Sachs K, Perez O, Peer D, Lauffenburger DA, Nolan GP. Causal protein-signaling networks derived from multiparameter single-cell data. Science. 2005;308: 523529.