Algorithm Comparison on Simulated Mental Health and Remote Work Data

Xinqi Huang xih037@ucsd.edu
Ruizhe Dai rdai@ucsd.edu
Vivian Zhao vxzhao@ucsd.edu
Yishan Cai yic075@ucsd.edu
Biwei Huang bih007@ucsd.edu
Jelena Bradic jbradic@ucsd.edu

Abstract

Recent shifts toward remote work due to global events such as the COVID-19 pandemic have catalyzed an urgent need to explore its effects on mental health. This project utilizes a simulated dataset designed to emulate the dynamics between mental health and remote work environments, specifically for the purpose of testing various causal inference algorithms. The simulated relation includes variables such as social isolation rating, average daily work hours, sleep quality, and company support for mental health. By employing advanced causal algorithms like PC with kernel conditional independence and Fisher-Z tests, Fast Causal Inference (FCI), and Greedy Equivalence Search (GES) integrated with a custom scoring method, this study seeks to determine which methods most accurately uncover the complex causal relationships. This approach aims to provide deeper insights into the structural dependencies and seeks to advance our understanding and application of these analytical tools in complex social science relations.

Code: https://github.com/VivianZhao12/CAPSTONE-Causal-Inference-between-Remote-Work-and-Mental-Health

1 Introduction . . . 2
2 Methods . . . 5
3 Results . . . 8
4 Discussion . . . 17
5 Conclusion . . . 18
6 Acknowledgments . . . 19

1 Introduction

1.1 Introduction

The generalization of remote work has significantly affected the dynamics of mental health, a trend that keeps on evolving with the ongoing normalization of remote work arrangements. Based on findings in the work of Bertoni et al. (2022), which explored the causal effects of remote working on the mental health of senior Europeans using longitudinal data from the SHARE survey during the initial COVID-19 outbreak, our project broadens this perspective by simulating a dataset that captures a more diverse demographic profile. This simulated approach allows us to rigorously test and compare the efficacy of various causal inference algorithms, including novel applications within the causal-learn library, to determine their effectiveness in different scenarios. Our analysis not only revisits established connections identified in prior research but also provides a critical evaluation of how different causal models perform in dissecting the links between remote work conditions and mental health outcomes.

1.2 Literature Review

Research indicates that remote work can significantly contribute to social isolation and technostress, ultimately affecting employee satisfaction and well-being. Variations in remote work satisfaction are evident across different age groups and living situations. Younger employees may experience increased isolation due to limited networking and mentorship opportunities crucial for career development, while older employees often face challenges with remote technology and team management (Tahlyan et al. 2024). Furthermore, individuals living alone tend to report heightened feelings of loneliness, emphasizing the connection between social interactions and overall mental health (Fingerman et al. 2020). The frequent use of information and communication technology (ICT) in remote settings can exacerbate technostress, leading to psychological strain and decreased job satisfaction (Khedhaouria et al. 2024). These findings underscore the necessity of examining the causal relationships between remote work dynamics and social isolation to better understand their implications for mental health and employee satisfaction.

In the realm of causal inference, the PC, FCI, and GES (Fingerman et al. 2020) are three classic and widely recognized algorithms used in clinical and public health research, contributing to advancements in precision medicine, pathophysiology, and social studies. The PC algorithm and its variants, such as PC-stable and conservative PC, address order dependence challenges, providing robust frameworks for uncovering causal relationships in complex datasets (Nogueira et al. 2022). These methods allow researchers to systematically analyze variable interactions without being biased by the order of assessments. For example, the PC algorithm has successfully elucidated causal relationships among phenotypes (Chaibub Neto et al. 2010) and in single-cell analyses (Wen et al. 2022).

While the PC and FCI algorithms are both constraint-based, beginning with a fully connected undirected graph and progressively removing edges between conditionally independent variables, the FCI algorithm notably extends the capabilities of PC by accommodating latent variables (Spirtes, Meek and Richardson 2013). This flexibility is crucial for analyzing real-world scenarios where unmeasured confounder can substantially influence causal conclusions. To further mitigate order dependence, the FCI algorithm can be executed multiple times with varying random orders of variables, enhancing the reliability of the resulting causal graph (Guo et al. 2024). This iterative approach helps ensure that identified causal structures are robust and reflective of the complex dynamics inherent in mental health impacts associated with remote work.

In contrast, GES is a score-based approach that starts with an empty graph and iteratively adds or removes edges in a greedy manner to optimize the fit measured by the chosen scoring method. Shen et al. (2020) conducted an extensive research on the application of the two methods FCI and FGES (a faster version of GES) in Alzheimer's pathophysiology, leading to their result that score-based FGES outperformed FCI in terms of precision and stability in various degrees of background knowledge (Shen et al. 2020). FCI was hindered by incorrect independence tests influenced by selection bias and data artifacts, which generated erroneous "V" or "Y" structures that propagated errors in the causal structure, however, FCI is capable of discovering a possible lagged relationship between (Shen et al. 2020).

1.3 Data Description

In our project, our objective is to compare various causal discovery and inference algorithms on a data set focused on remote work and mental wellness. Initially, we explored an online dataset, but its low correlations between variables made it unsuitable for our purposes. After discussing options with our mentor, we decided to simulate a dataset specifically structured around causal discovery.

In our simulated data set, we establish several key causal links. The social isolation rating directly affects the stress level, as increased isolation is often linked to increased stress due to a lack of social support. Family Number has a causal impact on Social Isolation Rating since individuals with larger families may experience less social isolation due to closer social ties. Education Level serves as a confounder between Average Daily Work Hours and Family Number. Education Level serves as the confounder because it independently influences both Average Daily Work Hours, often through career demands, and Family Number, through family planning choices associated with educational attainment. Additionally, Average Daily Work Hours influences Sleep, as longer working hours can reduce available time for rest, and Exercise Level has a direct effect on Sleep by helping improve sleep quality. Extroverted is a binary variable that indicate whether a person is extroverted or not. It serves as the confounder between Social Isolation Rating and Exercise Level because more extroverted person tends to have lower social isolation and also tend to do more exercise. Work Location impacts Exercise Level, as different environments can encourage or discourage physical activity, and Company Support for Remote Work affects Work Location by influencing whether employees work remotely or on-site.

We introduced nonlinear relationships to increase the complexity and provide a robust test for each algorithm. For instance, Average Daily Work Hours has a squared relationship with Sleep, meaning that as work hours increase, their negative impact on sleep intensifies; Family Number has a logarithmic relationship with Social Isolation Rating, reflecting that increases in family size provide diminishing social appeasement; and Social Isolation Rating has a sinusoidal relationship plus linear growth with Stress Level, capturing both the cyclical nature and steadily increasing impact of isolation on stress.

Here is our hypothesized causal map:

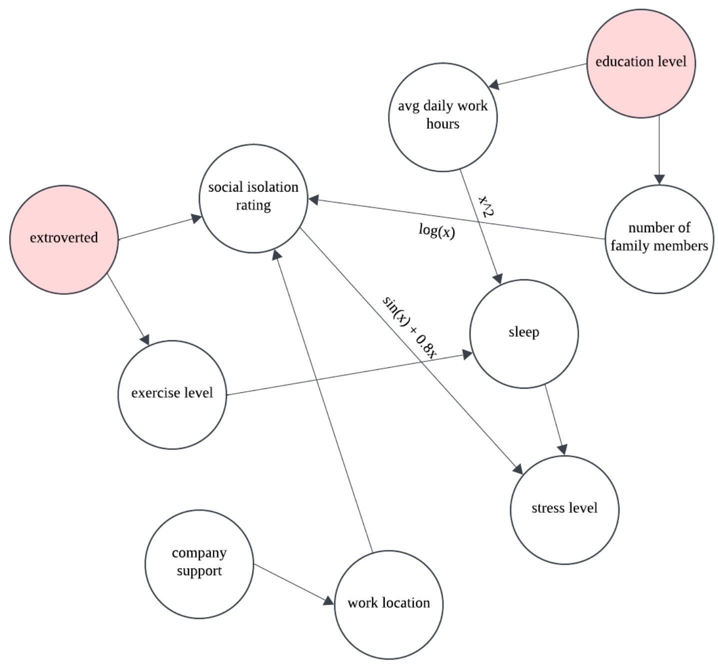


Figure 1: Hypothesized Ground Truth Causal Map of Remote Work, Social Isolation, and Wellness Dynamics

2 Methods

2.1 Data Simulation

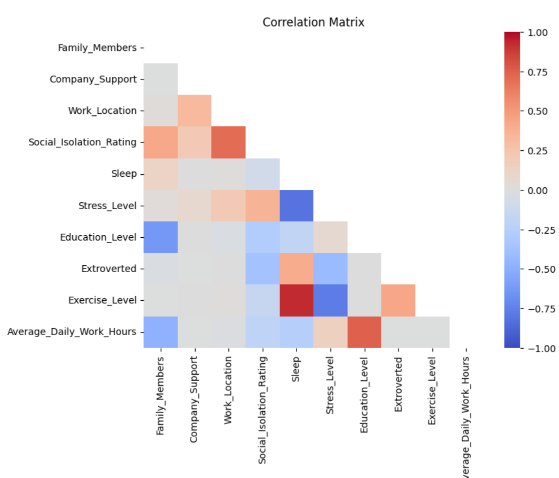
Variable Name	Simulation Method
education level	Randomly assigned from high_school: 1, bachelor: 2, master: 3, doctorate: 4, with uniform distribution.
avg daily work hours	8+education_level + Normal(0,1) with values clipped between 0-16
number of family members	2−education_level +Poisson(1) with values clipped between 0-10 and rounded to integer
extroverted	Randomly assigned from introverted: 0, extroverted: 1 with uniform distribution
exercise level	2+extroverted+Normal(0,1) with values clipped between 0 and 4
company support	Randomly assigned from 1-6 with uniform distribution
work location	$$p_{\text{remote}} = \frac{1}{1 + e^{-(\text{company\_support}-3)\cdot 0.5}}$$ work location = $$\begin{cases} \text{remote (0.6); hybrid (0.4)} & \text{if } U(0,1) < p_{\text{remote}}, \\ \text{onsite} & \text{otherwise.} \end{cases}$$
social isolation rating	$$\text{base} = \begin{cases} 3 + \mathcal{N}(0, 0.5) & if\,\text{remote}, \\ 2 + \mathcal{N}(0, 0.5) & if\,\text{hybrid}, \\ 1 + \mathcal{N}(0, 0.5) & if\,\text{onsite}. \end{cases}$$ social isolation rating = $$\text{base} + \log(1 + \text{number\_of\_family\_member}) - \text{extroverted}$$ with values clipped between 1-5
Sleep	$$2 - \left(\frac{\text{avg\_daily\_work\_hours}}{10}\right)^2 + \text{exercise\_level}$$ with values clipped between 1-10
stress level	$$\sin(\text{social\_isolation\_rating}) + 0.8 \cdot \text{social\_isolation\_rating} +$$ $$\text{(4 - sleep) + N(0, 0.5)}$$ with values clipped between 1-10

2.2 Exploratory Data Analysis

Figure 2 shows that the simulated dataset indeed aligned with our design.

- **Strong Positive Correlations between:**
  - *education level* and *average daily work hours*.
  - *work location* and *social isolation rating*.
  - *exercise level* and *sleep*.
- **Strong Negative Correlations between:**
  - *education level* and *number of family members*.
  - *sleep* and *stress level*.
  - *exercise level* and *stress level*.
- **Moderate Positive Correlations between:**
  - *number of family members* and *social isolation rating*.
  - *company support* and *work location*.
  - *social isolation rating* and *stress level*.
- **Moderate Negative Correlations:**
  - *average daily work hours* and *sleep*.
  - *extroverted* and *social isolation rating*.

While other weak correlations may result from causal chains, we observed a moderate positive correlation between *social isolation rating* and *number of family members*, which can be attributed to the logic used in simulating the data. Additionally, there is a moderate correlation between *work location* and *stress level*. However, these relationships represent correlations, not causation, and the underlying causal mechanisms require further investigation.



(a) Correlation Heatmap between simulated variables.

(b) Pairplot of the simulated variables.

Figure 2: EDA plots

2.3 Causal Discovery Algorithms

In this project, we have tested performances of three algorithms on the simulated data: PC algorithm, FCI algorithm, and the GES algorithm. In this section, we will generally introduce the working logic behind these algorithms.

2.3.1 PC Algorithm

The PC algorithm learns causal relationship in a dataset through iteratively testing conditional independence among variables. The algorithm starts on a initial fully connected graph, and then remove edges between variables that are independent with each other first without conditioning on other variables and then it will perform this action again conditioning on subsets of other variables. After this, the algorithm will move on to deciding the direction of remaining edges according to causal rule learned from independence/dependence pattern. This process results in a partially directed acyclic graph that represents possible causal relationships within the data.

2.3.2 FCI Algorithm

The FCI algorithm is an “advanced” version of the PC algorithm with the ability to detect confouding variables between two variables. It’s work flow is mostly similar with that of the PC algorithm, starting with a completely connected graph and gradually test conditional independeces and remove edges among variables. Nonetheless, it adopts a broader set of independence tests to identify patterns that suggest the possible of the presence of confounding variables. FCI outputs a Partial Ancestral Graph, which represents equivalence classes of causal structures consistent with the data, showing both definite and ambiguous causal relationships.

2.3.3 GES Algorithm

The GES algorithm takes a different approach to learn the causal relationship in a dataset. It uses a scoring criterion that it tries to optimize and also balance the model complexity and fit. In contrast to PC and FCI algorithms, the GES algorithm starts with an empty graph and gradually adds edges into the graph to improve its score, and it also has the greedy element in this process, which it always add the edge that gives it the most improvement. When the improvement it receives from adding additional edge is trivial, it will try to review the current graph and remove redundant edges(without lowering the score it already has) to make the graph simpler. In the end, the GES algorithm produces a Completed Partially Directed Acyclic Graph showing causal relationships among variables.

3 Results

3.1 PC Algorithm

3.1.1 Results with $\alpha = 0.05$

Fisher Z Test Comparing the PC algorithm result with the ground truth causal structure reveals both strengths and limitations of the algorithm's performance. The black edges in the graph indicate correctly identified causal relationships, demonstrating that the PC algorithm successfully captured many of the true causal connections in the network.

However, there are two notable discrepancies in the algorithm's output. First, the algorithm failed to identify the confounding relationship between *number of family members* and *average daily work hours*. In the ground truth, *education level* (shown in pink in the true graph) acts as a common cause for both variables. This omission highlights a key limitation of the PC algorithm in detecting latent confounders. Second, the algorithm incorrectly inferred a direct causal link between *work location* and *stress level* (indicated by the red edge). This spurious relationship suggests that the algorithm may be sensitive to indirect correlations in the data, potentially mistaking downstream effects for direct causal connections.

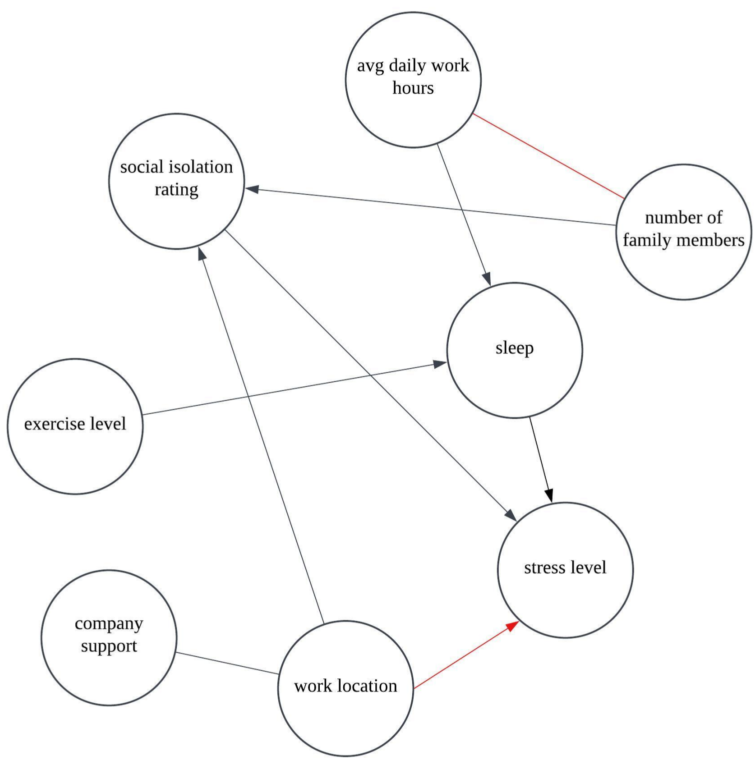


Figure 3: Causal Graph from PC Algorithm using Fisher Z Test ($\alpha = 0.05$)

KCI Test Similar to Fisher-Z, the PC algorithm did not identify the two pairs of confounding variables affecting both *average daily work hours* and *number of family members*, and *isolation rating* and *exercise level*. This is likely because the KCI test, while capable of detecting non-linear dependencies, may struggle with complex dependencies masked by confounding variables. In our simulation, these two variables are influenced by *education level*, which affects both variables but was not identified as a confounder by the PC algorithm.

The omission of the causal edge from *sleep* to *stress level* could be attributed to the inability of the KCI test to capture the true dependency pattern, possibly due to the complex interaction defined by your simulation equations. The relationship involves non-linear transformations and interactions with other variables, which might not have been effectively isolated during the conditional independence testing phase. Additionally, the manipulation of the sleep variable's effect on stress, combined with noise addition, might have attenuated the detectable strength of the causal effect to below the threshold for detection at the chosen significance level.

The failure to correctly identify the directionality of the causal relationship between *company support* and *work location* likely stems from the conditional mechanism used in the simulation, which defines *work location* as a function of *company support*.

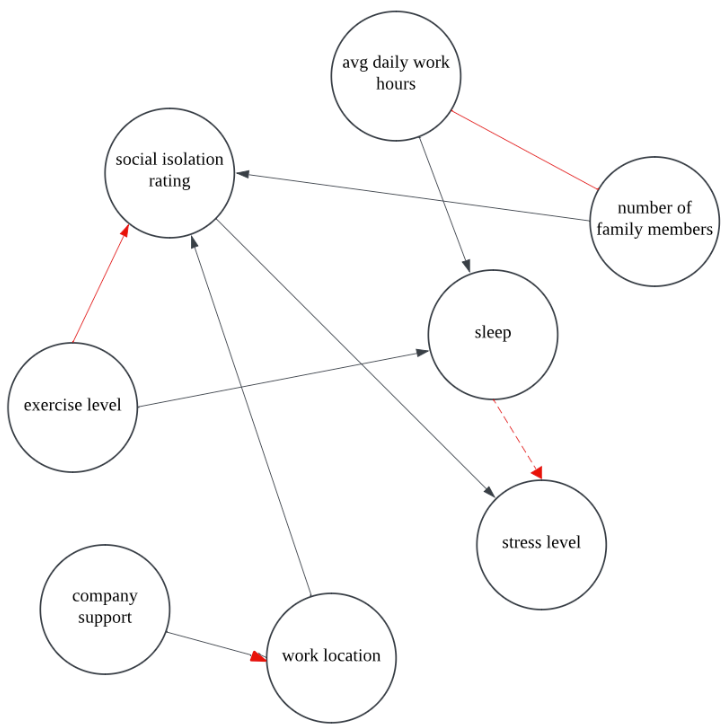


Figure 4: Causal Graph from PC Algorithm using KCI Test ($\alpha = 0.05$)

3.1.2 Results with $\alpha = 0.01$

Fisher Z Test Fisher-Z assesses linear correlations between variables and typically assumes that the data are normally distributed, which may not hold true in complex simulation scenarios. The presence of a confounding variable, such as *education level* affecting both these variables, can introduce correlations that Fisher-Z might interpret as direct causal links due to its focus on linear associations.

Similar to the first issue, if *social isolation rating* and *exercise level* are influenced by an unaccounted confounder, the Fisher-Z test may again be inadequate. The test's linear framework can misinterpret underlying dependencies influenced by confounders as direct causal relationships, especially when these dependencies are non-linear or when the confounder's effects are complex and not clearly delineated through linear measures.

The test in correctly determining the causal direction between *company support* and *work location* likely stems from the probabilistic or logistic nature of how *work location* might depend on *company support*, as indicated in your simulation parameters. The Fisher-Z test, focusing predominantly on correlation, is less equipped to deduce causal direction in scenarios where the relationship is governed by non-linear or threshold-based mechanisms. This leads to difficulties in asserting directional influence when the relationship between the variables is contingent upon specific conditions or thresholds, as might be dictated by logistic functions.

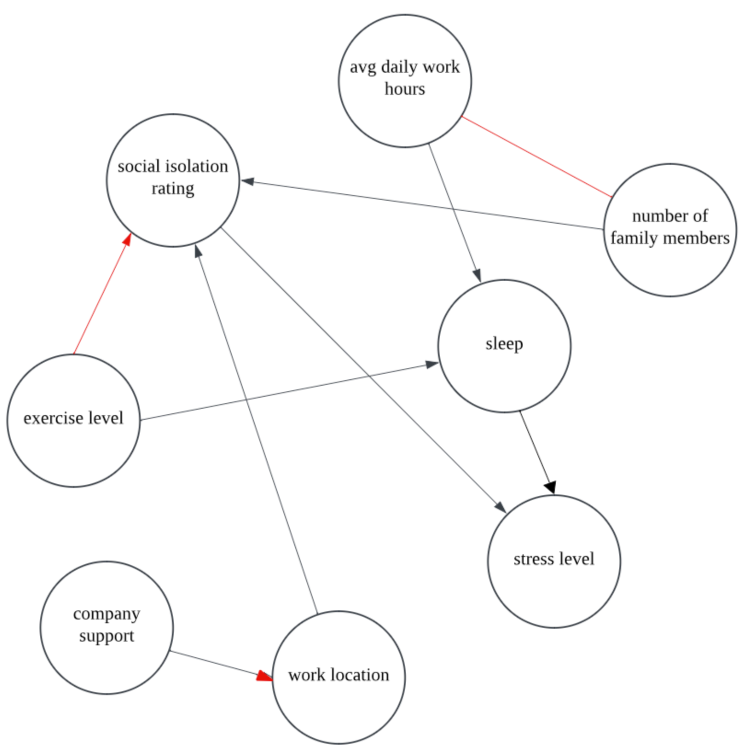


Figure 5: Causal Graph from PC Algorithm using Fisher-Z Test ($\alpha = 0.01$)

KCI Test The PC+KCI approach did not identify the hidden confounder that affects both *average daily work hours* and *number of family members*. KCI, although capable of detecting non-linear dependencies, can still miss confounders if the indirect effects or the nature of the dependency does not sufficiently impact the joint distribution in a way that is detectable at the chosen significance level.

The missing causal link between *sleep* and *stress level* is likely due to the non-linear and complex nature of their relationship as described in your simulation. KCI tests are based on kernel methods that handle non-linearity by transforming the data into a high-dimensional space where linear relationships are assessed. However, if the kernel used does not align well with the specific type of non-linearity or if the noise in the data obscures the relationship, the KCI test may fail to detect it. Again, the strict significance level used (0.01) could also contribute to this issue by requiring a very high degree of certainty to confirm dependencies, potentially overlooking genuine causal effects.

The incorrect or unrecognized directional relationship between *company support* and *work location* can be a result of several factors. The issue here could stem from the non-linear nature of this dependency (modeled via a logistic function) not being captured correctly by the kernel function used in the KCI test. Additionally, if the data does not exhibit enough variability in *company support* to clearly influence the distribution of *work location*, the algorithm might fail to establish the correct causal direction.

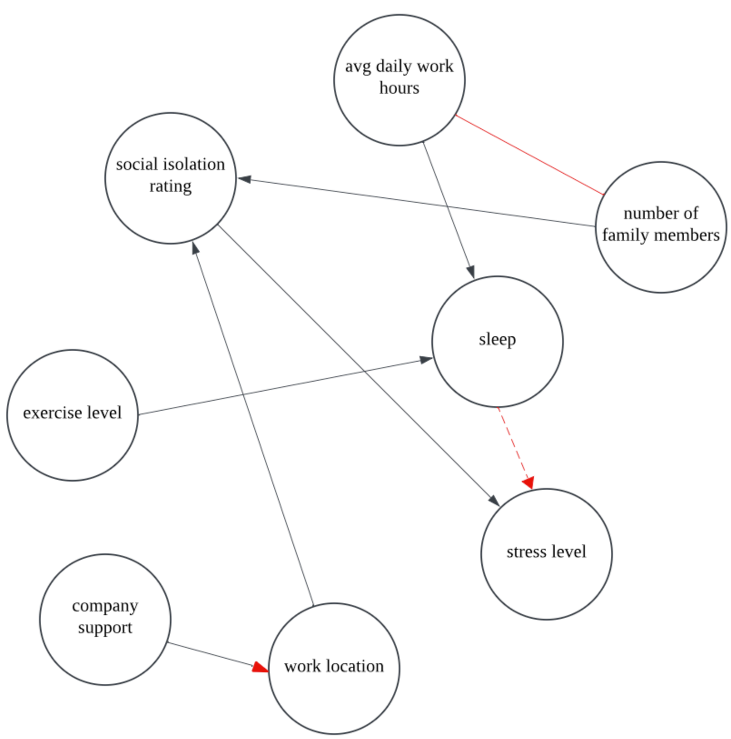


Figure 6: Causal Graph from PC Algorithm using KCI Test ($\alpha = 0.01$)

3.2 FCI Algorithms

3.2.1 Fisherz and FastKCI Independence Tests

The FCI algorithm with Fisher Z and FastKCI tests at $\alpha = 0.05$ and at $\alpha = 0.01$ offer the same causal relationship graph. They correctly identified many direct causal relationships shown in black edges as PC algorithm did. However, they incorrectly inferred a direct causal link between *average daily work hours* and *number of family members* shown in red, and erroneously suggested a bidirected edge between *social isolation rating* and *stress level*. This bidirected edge implies a latent common cause between these variables, which deviates from the ground truth structure where there is actually a direct causal relationship from *social isolation rating* to *stress level*. Notably, both Fisher Z and Fast KCI tests produced identical results at these significance levels. While the Fisher Z test assumes linear relationships and normal distribution, and FastKCI can detect non-linear dependencies, this similarity in results indicates that the relationships between variables in our simulated dataset might be predominantly linear in nature. This observation is supported by our data generation process where most of the relationships were simulated using linear or simple non-linear functions (like x^2 and $log(x)$), which both tests were able to capture equally well at the 0.05 and 0.01 significance levels.

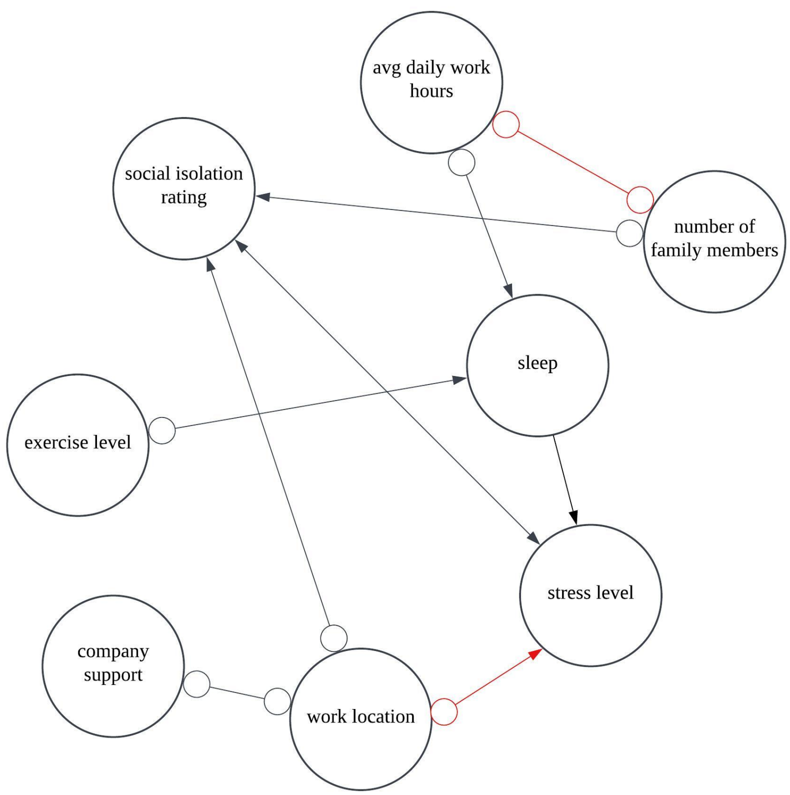


Figure 7: Causal Graph from FCI Algorithm using Fisher Z and FastKCI Tests ($\alpha = 0.05$ and $\alpha = 0.01$)

3.2.2 KCI Independence Tests

$\alpha = 0.05$ The algorithm produced three deviations from the ground truth structure. While many relationships were correctly identified, a spurious direct causal link emerged between *average daily work hours* and *number of family members* shown in solid red, despite these variables being connected only through the latent variable *education level* in reality. Among its shortcomings, the algorithm also missed the existing causal relationship between *sleep* and *stress level*, as indicated by the dashed red line. Perhaps most notably, the results show an incorrectly oriented causal flow from *exercise level* to *social isolation rating* marked in solid red, overlooking their true connection through the latent variable *extroverted*.

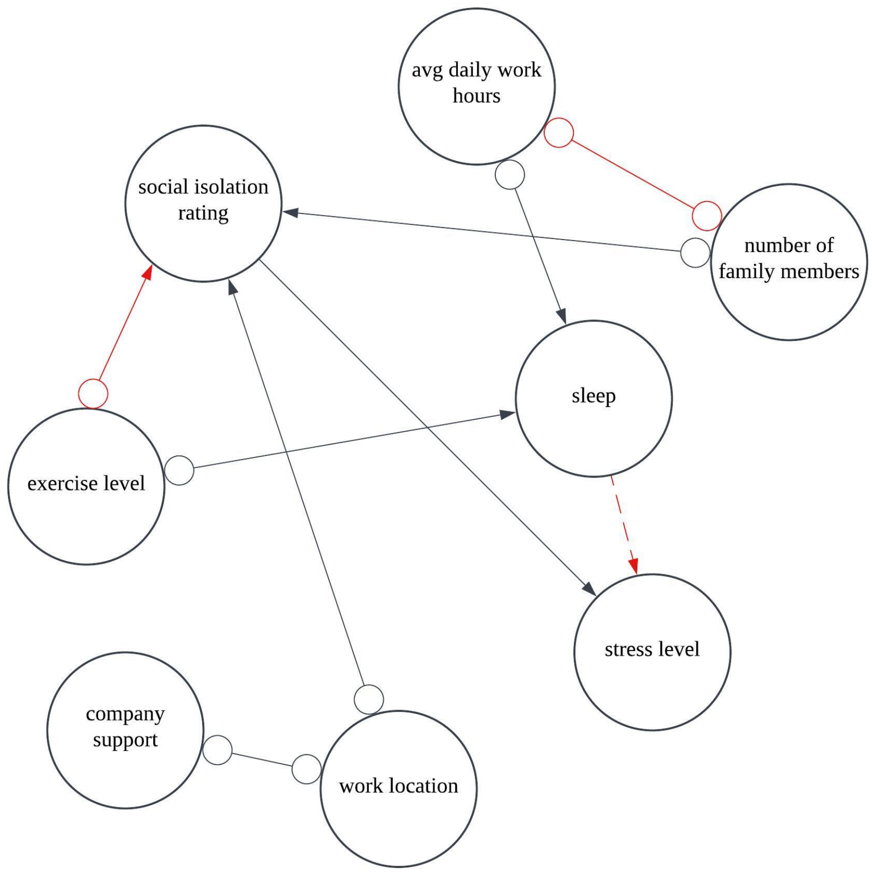


Figure 8: Causal Graph from FCI Algorithm using KCI Test ($\alpha = 0.05$)

$\alpha = 0.01$ When decreasing the significance level to $\alpha = 0.01$, the algorithm shows some improvement but still maintains certain errors. While it corrects the direction of the relationship between *exercise level* and *social isolation rating*, it continues to incorrectly infer a direct causal link between *average daily work hours* and *number of family members* shown in solid red and fails to detect the causal relationship between *sleep* and *stress level* indicated by the dashed red line. Notably, the computational cost of the KCI test is substantial, requiring over 420 minutes to complete the analysis, which indicates significant computational overhead compared to other independence tests.

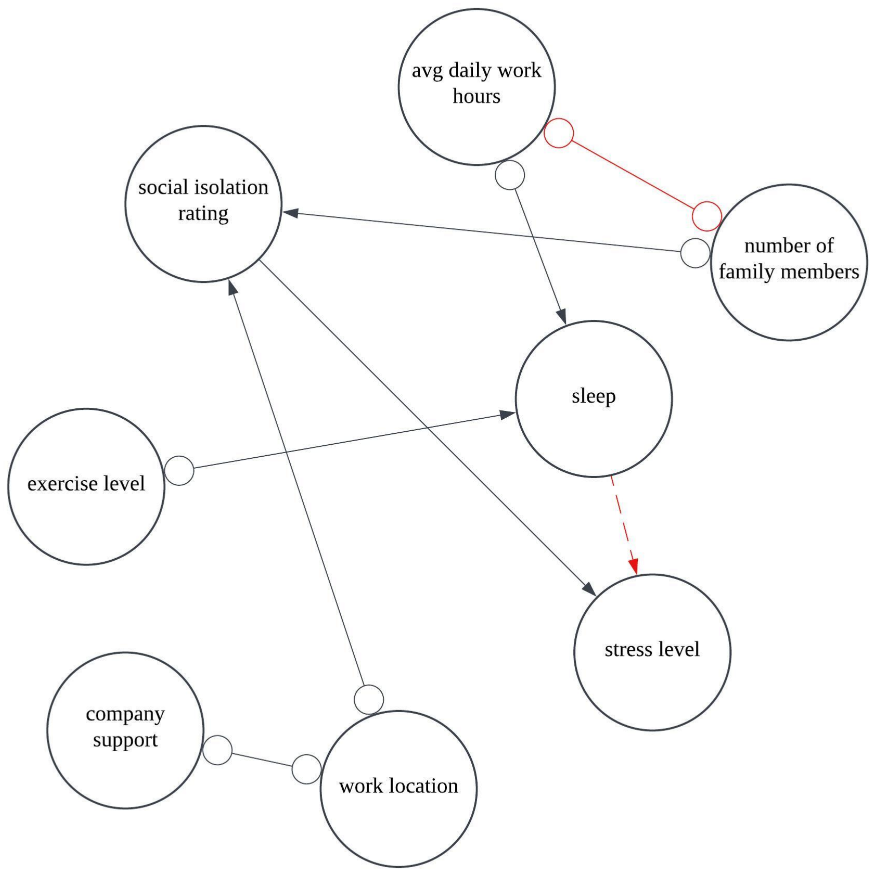


Figure 9: Causal Graph from FCI Algorithm using KCI Test ($\alpha = 0.01$)

3.3 GES Algorithm

GES with BIC score Compared to constraint-based methods, GES correctly identified all directed causal relationships represented by black edges, demonstrating it's robustness against both linear and non-linear relationships. However, we observed three discrepancies between the resulting graph and the ground truth. Notably, GES incorrectly inferred a direct causal link between *work location* and *stress level* (marked as red in Figure 8), indicating it's sensitivity to indirect correlations in the simulated data, similar to the PC algorithm. Additionally, GES failed to account for the effects of the latent confounders *education level* and *extroverted*. As shown in the graph, we observed a directed edge between *exercise level* and *social isolation rating,* and an undirected edge between *average daily work hours* and *number of family members*. This highlights GES's limitations in handling latent variables, as it may introduce spurious edges due to its inability to represent or account for them. The potential reason for this can be traced back to GES's greedy nature, where the algorithm maximizes the scoring criterion, favoring locally optimal solutions while failing to penalize models for spurious correlations induced by unmeasured confounders.

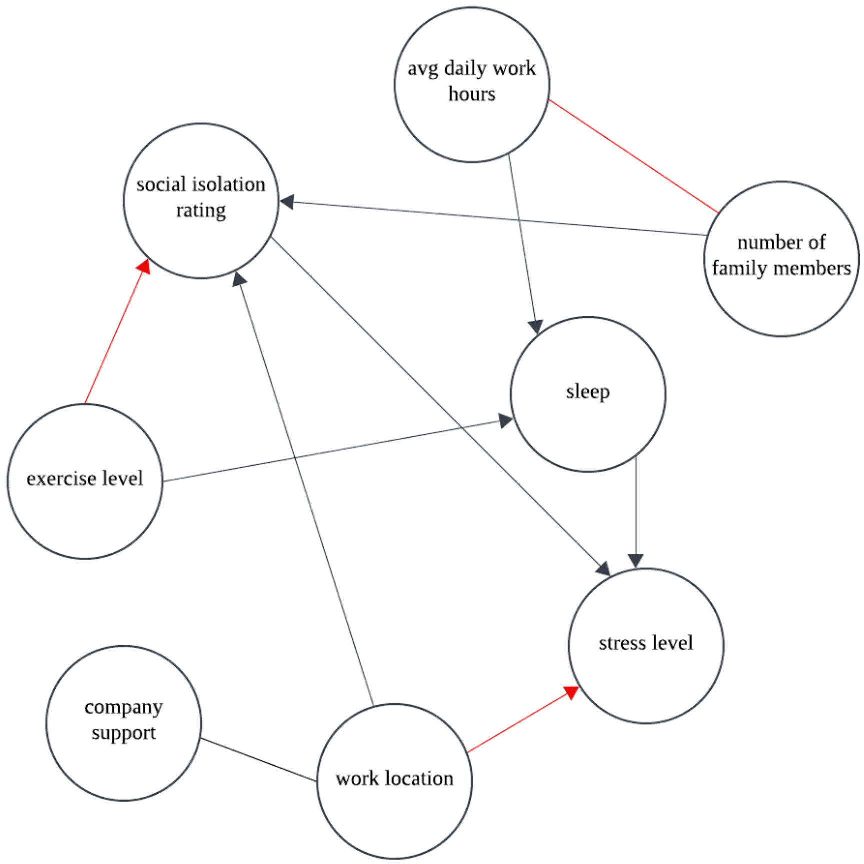


Figure 10: Causal Graph from GES Algorithm using BIC score

3.4 ANM Analysis

Given the discrepancies in causal relationships identified by our algorithms, we employed Additive Noise Models (ANM) for further investigation. ANM is a framework that determines the direction of cause-effect relationships by exploiting asymmetries in the data generation process. In ANM, an effect Y is modeled as a function of its cause X plus independent noise: $Y = f(X) + N$, where f is a potentially non-linear function and N is a noise term independent of X. This asymmetric relationship helps infer causal direction.

We specifically examined two questionable relationships: the link between *family members* and *average daily work hours*, and the relationship between *social isolation rating* and *exercise level*. For the first pair, initial ANM analysis suggested a causal direction from *family members* to *average daily work hours* (p-value = 3.45e-06). However, after conditioning on *education level*, both directional p-values became non-significant (0.996 and 0.995), confirming our understanding that there is no direct causal relationship between these variables when accounting for the latent confounder.

For the second pair, initial ANM analysis indicated a bidirectional causal relationship between *social isolation rating* and *exercise level*, with both directions showing statistical significance. After conditioning on *extroverted*, both directions still showed strong dependency (p-values = 1.41e-07). This unexpected result led us to consider additional confounding factors. When we extended our conditioning to include multiple predictors (*extroverted*, *work location*, and *family members*), the p-values became non-significant (0.117 and 0.125), suggesting no direct causal relationship in either direction after accounting for these combined influences. This analysis demonstrates the importance of considering multiple confounding factors in causal discovery and validates our ground truth structure where these variables are connected through the latent variable *extroverted*.

3.5 Evaluation

To ensure fair comparison between different causal discovery algorithms, we established a standardized evaluation process. For the PC algorithm outputs, we converted undirected edges to directional edges by comparing with the ground truth graph direction. For FCI and GES algorithm outputs, we transformed circle endpoints to either dashes or arrowheads based on the closest possible match to the ground truth graph. All graph representations were then converted to GeneralGraph objects using the txt2generalgraph method for consistent comparison.

Using these standardized graph representations, we computed several evaluation metrics. The Structural Hamming Distance (SHD) measures the minimum number of edge additions, deletions, or reorientations needed to transform the predicted graph into the true graph, with lower values indicating better performance. We also computed two types of F1 scores that balance precision and recall. The Arrow F1 score evaluates the accuracy of both the presence and direction of causal relationships, while the Adjacency F1 score focuses solely on the presence of connections between variables, regardless of their direction. The runtime measurement provides practical insights into computational efficiency.

For example, in PC algorithm with fisherz and 0.05 significance level, an SHD of 2 indicates that two edge operations would be needed to match the ground truth. The Arrow F1 score of 87.50% suggests good accuracy in identifying correct causal directions, while the higher Adjacency F1 score of 87.50% shows that the algorithm was particularly successful in identifying the presence of relationships between variables, even if some directional assignments were incorrect. These metrics together provide a comprehensive view of each algorithm's strengths and limitations in recovering the true causal structure.

Table 1: Performance Comparison of Causal Discovery Algorithms

Algorithm	Configuration	SHD	Arrow F1(%)	Adjacency F1(%)
PC	Fisher Z ($\alpha = 0.05$)	2	87.50	87.50
	Fisher Z ($\alpha = 0.01$)	2	87.50	87.50
	KCI ($\alpha = 0.05$)	3	80.00	80.00
	KCI ($\alpha = 0.01$)	2	85.71	85.71
FCI	Fisher Z ($\alpha = 0.05$)	3	77.78	87.50
	Fisher Z ($\alpha = 0.01$)	3	77.78	87.50
	Fast KCI ($\alpha = 0.05$)	3	77.78	87.50
	Fast KCI ($\alpha = 0.01$)	3	77.78	87.50
	KCI ($\alpha = 0.05$)	3	70.59	80.00
	KCI ($\alpha = 0.01$)	2	80.00	85.71
GES	BIC score	3	82.35	82.35

4 Discussion

4.1 Results

The evaluation of various causal discovery algorithms as detailed in Table 1 showcases the relative strengths and weaknesses of each algorithm across different configurations and criteria such as Structural Hamming Distance (SHD), Arrow F1 score, and Adjacency F1 score. The PC algorithm, when tested with the Fisher Z test, consistently shows strong performance with an SHD of 2 and high F1 scores both in Arrow and Adjacency across both significance levels ($\alpha = 0.05$ and $\alpha = 0.01$). This suggests a robust ability to correctly identify and orient edges, making it a reliable choice for scenarios where the data approximates the assumptions of linear relationships and normal distributions. The KCI variant of the PC algorithm, while showing slightly less robustness at a significance level of 0.05, improves significantly when the stringency is increased to $\alpha = 0.01$, suggesting its enhanced capability to deal with non-linear dependencies under stricter testing conditions.

The FCI algorithm's consistent performance under Fisher Z and FastKCI independence tests suggests that the relationship between variables in our simulated dataset is clear and well-defined. Moreover, these structured relationships have moderate noise that makes causal signal detectable even at stricter significance level. The improvement of FCI's performances under the KCI independence test from significan level 0.05 to 0.01 indicates the inherent difference of KCI with Fisher Z and FastKCI, which KCI is more prone to allow weak causal signals to pass as significant at looser threshold, whereas Fisher Z and FastKCI are less prone to this issue.

Lastly, the GES algorithm, employing a score-based approach and beginning its analysis with an empty graph, shows moderate success with an SHD of 3 and an Arrow and Adjacency F1 of 82.35%. This performance suggests that while GES is quite effective in identifying causal relationships, it might introduce errors in the presence of latent variables or complex dependency structures. Its strategy of iteratively adding edges based on a scoring system seems to be less effective at avoiding spurious correlations compared to PC and FCI, which systematically test for conditional independencies.

4.2 Limitations

The reason of using simulated dataset is because real world dataset is too noisy and not feasible for us to clean the dataset and perform causal discovery algorithm trial and comparisons on that real world dataset in the first quarter. As we discussed with our mentor, we decided to use simulated dataset for Quarter 1's project and use real world dataset in Quarter 2. However, there are limitations in out quarter 1 project: the simulation relies heavily on predefined assumptions, which might oversimplify the complexities of real world scenarios and may not be as representative as using a real-world dataset. Additionally, the use of specific probability distribution, such as the normal and the poisson distribution may not accurately reflect that of the real world cases, which may limit our result's generalizability to complex scenarios.

5 Conclusion

This project extensively compared the effectiveness of several causal inference algorithms, including PC, FCI, and GES, in accurately modeling a simulated dataset reflecting the dynamics of remote work and mental health. Our results illuminate the varied capabilities of these algorithms in replicating the ground truth causal structure we designed. The PC algorithm demonstrated robustness in capturing many true causal relationships, yet it occasionally failed to identify and correctly interpret confounding variables and the direction of causality in complex setups. Notably, the FCI algorithm, while similar in approach to PC, offered slight improvements by incorporating considerations for latent variables, but still showed vulnerabilities in fully deciphering the underlying causal mechanisms, especially in the presence of non-linear relationships.

The GES algorithm, using a score-based approach, differed fundamentally from PC and FCI by starting with an empty graph and iteratively adding edges. It proved effective in identifying direct causal links and demonstrated a higher resilience against misinterpreting indirect relationships as direct causalities. However, its performance also indicated limitations in detecting and accounting for latent confounders, highlighting a common challenge across all tested algorithms. These findings underscore the critical need for algorithm refinement and the development of more sophisticated methods that can more accurately decode complex causal structures, particularly when latent variables and non-linear dependencies are involved.

Moving forward, this study not only provides a benchmark for comparing the efficacy of well-known causal inference algorithms but also sets the stage for future research aimed at enhancing these methods. As we continue to refine these computational tools, it is imperative to integrate more adaptive and nuanced techniques capable of handling the complexities inherent in real-world data. This will ultimately lead to more accurate and actionable insights in various domains, from public health to social sciences, where understanding the nuances of causality is crucial.

6 Acknowledgments

In our final report on the algorithmic comparison for simulated mental health and remote work data, we convey our deepest appreciation to our mentors, Biwei Huang and Jelena Bradic, for their expert guidance and steadfast support. Their unwavering commitment to this research has been instrumental in our understanding and application of advanced causal inference techniques. We are grateful to the University of California, San Diego for providing us the opportunity to conduct this meaningful study. We are deeply thankful for the collective knowledge and collaborative spirit of everyone involved in this project.

References

Bertoni, Marco, Danilo Cavapozzi, Giacomo Pasini, and Caterina Pavese. 2022. "Remote Working and Mental Health During the First Wave of the COVID-19 Pandemic." [Link]

Chaibub Neto, Elias, Mark P. Keller, Alan D. Attie, and Brian S. Yandell. 2010. "Causal graphical models in systems genetics: A unified framework for joint inference of causal network and genetic architecture for correlated phenotypes." [Link]

Fingerman, Karen L, Yee To Ng, Shiyang Zhang, Katherine Britt, Gianna Colera, Kira S Birditt, and Susan T Charles. 2020. "Living Alone During COVID-19: Social Contact and Emotional Well-being Among Older Adults." [Link]

Guo, Xinyu, Linna Wang, Zhenchao Li, Ziliang Feng, Li Lu, Lihua Jiang, and Li Zhao. 2024. "Factors and pathways of non-suicidal self-injury in children: insights from computational causal analysis." [Link]

Khedhaouria, Anis, Francesco Montani, Arshad Jamal, and Mahmood Hussain Shah. 2024. "Consequences of technostress for users in remote (home) work contexts during a time of crisis: The buffering role of emotional social support." [Link]

Nogueira, Ana Rita, Andrea Pugnana, Salvatore Ruggieri, Dino Pedreschi, and João Gama. 2022. "Methods and tools for causal discovery and causal inference." [Link]

Shen, Xinpeng, Sisi Ma, Prashanthi Vemuri, Gyorgy Simon, the Alzheimer's Disease Neuroimaging Initiative, Michael W. Weiner, Paul Aisen, Ronald Petersen, Clifford R. Jack, Andrew J. Saykin, William Jagust, John Q. Trojanowki, Arthur W. Toga, Laurel Beckett, Robert C. Green, John Morris, Leslie M. Shaw, Zaven Khachaturian, Greg Sorensen, Maria Carrillo, Lew Kuller, Marc Raichle, Steven Paul, Peter Davies, Howard Fillit, Franz Hefti, David Holtzman, M. Marcel Mesulam, William Potter, Peter Snyder, Adam Schwartz, Tom Montine, Ronald G. Thomas, Michael Donohue, Sarah Walter, Devon Gessert, Tamie Sather, Gus Jiminez, Archana B. Balasubramanian, Jennifer Mason, Iris Sim, Danielle Harvey, Matthew Bernstein, Nick Fox, Paul Thompson, Norbert Schuff, Charles DeCArli, Bret Borowski, Jeff Gunter, Matt Senjem, David Jones, Kejal Kantarci, Chad Ward, Robert A. Koeppe, Norm Foster, Eric M. Reiman, Kewei Chen, Chet Mathis, Susan Landau, Nigel J. Cairns, Erin Franklin, Lisa Taylor-Reinwald, Virginia Lee, Magdalena Korecka, Michal Figurski, Karen Crawford, Scott Neu, Tatiana M. Foroud, Steven Potkin, Kelley Faber, Sungeun Kim, Kwangsik Nho, Leon Thal, Neil Buckholtz, Marilyn Albert, Richard Frank, John Hsiao, Jeffrey Kaye, Joseph Quinn, Lisa Silbert, Betty Lind, Raina Carter, Sara Dolen, Lon S. Schneider, Sonia Pawluczyk, Mauricio Beccera, Liberty Teodoro, Bryan M. Spann, James Brewer, Helen Vanderswag, Adam Fleisher, Judith L. Heidebrink, Joanne L. Lord, Sara S. Mason, Colleen S. Albers, David Knopman, Kris Johnson, Rachelle S. Doody, Javier Villanueva-Meyer, Valory Pavlik, Victoria Shibley, Munir Chowdhury, Susan Rountree, Mimi Dang, Yaakov Stern, Lawrence S. Honig, Karen L. Bell, Beau Ances, Maria Carroll, Mary L. Creech, Erin Franklin, Mark A. Mintun, Stacy Schneider, Angela Oliver, Daniel Marson, David Geldmacher, Marissa Natelson Love, Randall Griffith, David Clark, John Brockington, Erik Roberson, Hillel Grossman, Effie Mitsis, Raj C. Shah, Leyla deToledo Morrell, Ranjan Duara, Maria T. Greig-Custo, Warren Barker, Chiadi Onyike, Daniel D'Agostino, Stephanie Kielb, Martin Sadowski, Mohammed O. Sheikh, Anaztasia

Ulysse, Mrunalini Gaikwad, P. Murali Doraiswamy, Jeffrey R. Petrella, Salvador Borges-Neto, Terence Z. Wong, Edward Coleman, Steven E. Arnold, Jason H. Karlawish, David A. Wolk, Christopher M. Clark, Charles D. Smith, Greg Jicha, Peter Hardy, Partha Sinha, Elizabeth Oates, Gary Conrad, Oscar L. Lopez, Mary Ann Oakley, Donna M. Simpson, Anton P. Porsteinsson, Bonnie S. Goldstein, Kim Martin, Kelly M. Makino, M. Saleem Ismail, Connie Brand, Adrian Preda, Dana Nguyen, Kyle Womack, Dana Mathews, Mary Quiceno, Allan I. Levey, James J. Lah, Janet S. Cellar, Jeffrey M. Burns, Russell H. Swerdlow, William M. Brooks, Liana Apostolova, Kathleen Tingus, Ellen Woo, Daniel H. S. Silverman, Po H. Lu, George Bartzokis, Neill R Graff-Radford, Francine Parfitt, Kim Poki-Walker, Martin R. Farlow, Ann Marie Hake, Brandy R. Matthews, Jared R. Brosch, Scott Herring, Christopher H. Van Dyck, Richard E. Carson, Martha G. MacAvoy, Pradeep Varma, Howard Chertkow, Howard Bergman, Chris Hosein, Sandra Black, Bojana Stefanovic, Curtis Caldwell, Ging-Yuek Robin Hsiung, Benita Mudge, Vesna Sossi, Howard Feldman, Michele Assaly, Elizabeth Finger, Stephen Pasternack, Irina Rachisky, John Rogers, Dick Trost, Andrew Kertesz, Charles Bernick, Donna Munic, Emily Rogalski, Kristine Lipowski, Sandra Weintraub, Borna Bonakdarpour, Diana Kerwin, Chuang-Kuo Wu, Nancy Johnson, Carl Sadowsky, Teresa Villena, Raymond Scott Turner, Kathleen Johnson, Brigid Reynolds, Reisa A. Sperling, Keith A. Johnson, Gad Marshall, Jerome Yesavage, Joy L. Taylor, Barton Lane, Allyson Rosen, Jared Tinklenberg, Marwan N. Sabbagh, Christine M. Belden, Sandra A. Jacobson, Sherye A. Sirrel, Neil Kowall, Ronald Killiany, Andrew E. Budson, Alexander Norbash, Patricia Lynn Johnson, Thomas O. Obisesan, Saba Wolday, Joanne Allard, Alan Lerner, Paula Ogrocki, Curtis Tatsuoka, Parianne Fatica, Evan Fletcher, Pauline Maillard, John Olichney, Charles DeCarli, Owen Carmichael, Smita Kittur, Michael Borrie, T.-Y. Lee, Rob Bartha, Sterling Johnson, Sanjay Asthana, Cynthia M. Carlsson, Pierre Tariot, Anna Burke, Ann Marie Milliken, Nadira Trncic, Adam Fleisher, Stephanie Reeder, Vernice Bates, Horacio Capote, Michelle Rainka, Douglas W. Scharre, Maria Kataki, Brendan Kelly, Earl A. Zimmerman, Dzintra Celmins, Alice D. Brown, Godfrey D. Pearlson, Karen Blank, Karen Anderson, Laura A. Flashman, Marc Seltzer, Mary L. Hynes, Robert B. Santulli, Kaycee M. Sink, Leslie Gordineer, Jeff D. Williamson, Pradeep Garg, Franklin Watkins, Brian R. Ott, Geoffrey Tremont, Lori A. Daiello, Stephen Salloway, Paul Malloy, Stephen Correia, Howard J. Rosen, Bruce L. Miller, David Perry, Jacobo Mintzer, Kenneth Spicer, David Bachman, Nunzio Pomara, Raymundo Hernando, Antero Sarrael, Susan K. Schultz, Karen Ekstam Smith, Hristina Koleva, Ki Won Nam, Hyungsub Shim, Norman Relkin, Gloria Chaing, Michael Lin, Lisa Ravdin, Amanda Smith, Balebail Ashok Raj, and Kristin Fargher. 2020. "Challenges and Opportunities with Causal Discovery Algorithms: Application to Alzheimer's Pathophysiology." [Link]

Spirtes, Peter L., Christopher Meek, and Thomas S. Richardson. 2013. "Causal Inference in the Presence of Latent Variables and Selection Bias." arXiv. [Link]

Tahlyan, Divyakant, Hani Mahmassani, Amanda Stathopoulos, Maher Said, Susan Shaheen, Joan Walker, and Breton Johnson. 2024. "In-person, hybrid or remote? Employers' perspectives on the future of work post-pandemic." [Link]

Wen, Yujian, Jielong Huang, Hai Zhang, Shuhui Guo, Yehezqel Elyahu, Alon Mon-

sonego, Yanqing Ding, and Hao Zhu. 2022. "CausalCell: applying causal discovery to single-cell analyses." Cold Spring Harbor Laboratory. [Link]