Exploring the Causal Impacts of Student Health on Online Learning Experiences During the Pandemic

Catherine Yeh cy3@williams.edu
Computer Science Department
Williams College

1. Introduction

COVID-19 has drastically impacted all of our lives, but one group of individuals hit particularly hard by the pandemic was students. As learning abruptly transitioned from in-person to online, students were faced with countless unknowns and novel difficulties as they navigated this brand new world of education. These problems were only compounded by health issues, social isolation, and other challenges induced by major societal shifts. Moving forward, we may see more forms of remote/hybrid learning emerge and persist. To better support students in these novel classroom settings, we must gain a better understanding of how students' lives have been impacted by the pandemic. This knowledge can help inform measures that should be taken to mitigate the negative repercussions of online education and craft more accommodating learning environments.

To explore the impact of COVID-19 in education, I used Chaturvedi et al. (2021)'s data set, which contains survey data from students in Delhi National Capital Region (NCR), India during the pandemic. In particular, I wanted to investigate the following research question: *How does health impact students' online class experiences?* To address this question, I began by performing data pre-processing and cleaning. Then, I followed the standard causal inference pipeline, starting by eliciting a graph over the variables of interest using causal discovery and background knowledge. Next, I used the backdoor criterion to identify the causal effect as a function of the observed data. Finally, I computed the average causal effect of health on online class experience using augmented inverse probability weighting.

My analysis revealed that under faithfulness, linearity, and edge presence assumptions, students with health issues tended to have more negative online class experiences than their peers. This suggests that improving student health considerations and interventions in online learning may lead to more effective forms of remote education.

2. Preliminaries

A directed acyclic graph (DAG) $\mathcal{G}$ is defined to be a set of vertices V that contains only directed edges ($\rightarrow$) such that there are no directed cycles (i.e., for all $V_i \in V$, there is no sequence of edges in $\mathcal{G}$ such that $V_i \rightarrow \ldots \rightarrow V_i$).

Statistical models of a DAG $\mathcal{G}$ are sets of distributions that factorize as: $p(V) = \prod_{V_i \in V} p(V_i \mid \text{pa}_{\mathcal{G}}(V_i))$, where $\text{pa}_{\mathcal{G}}(V_i)$ denotes the parents of V_i in $\mathcal{G}$. Conditional independences in $p(V)$ can be read off from $\mathcal{G}$ via d-separation (Pearl, 2009). That is, for disjoint sets X, Y, Z of V, we have that the following *global Markov property* holds: $(X \perp\!\!\!\perp Y \mid Z)_{\text{d-sep}} \implies (X \perp\!\!\!\perp Y \mid Z)_{\text{in } p(V)}$.

Causal DAG models can be represented as a triple containing the DAG $\mathcal{G}$ itself, a system of non-parametric structural equations with independent errors equipped with the do-operator, and the resulting set of distributions that factorize according to $\mathcal{G}$ as described above. Under this interpretation, a directed edge $V_i \rightarrow V_j$ can be interpreted to mean that V_i is potentially a direct cause of V_j.

Given a DAG $\mathcal{G}$, we can construct a Single World Intervention Graph (SWIG) $\mathcal{G}(a)$ that directly encodes conditional independences between factuals and counterfactuals associated with a specific hypothetical intervention on the treatment A (Richardson and Robins, 2013). This effectively involves executing the graphical operator $\text{do}(A = a)$ while allowing for the non-intervened version of the treatment to remain on the graph. The explicit construction of a SWIG involves first copying over G into $\mathcal{G}(a)$. Then, we split A into two vertices: a "random" vertex A and a "fixed" vertex a. A inherits all incoming edges in the initial DAG G, and a inherits all outgoing edges. Finally, we convert all descendants of a in $\mathcal{G}(a)$ (e.g, Y) into potential outcomes (e.g., $Y(a)$). An example SWIG is illustrated in Figure 2.

2.1 Assumptions

To facilitate structure learning, I will restrict my analysis to the set of *faithful* distributions with respect to a DAG $\mathcal{G}$ where $(X \perp\!\!\!\perp Y \mid Z)_{\text{d-sep}} \iff (X \perp\!\!\!\perp Y \mid Z)_{\text{in } p(V)}$. In addition, I make the simplifying assumption that the relations between my variables are linear, excluding the sensitivity analysis portion of my project (Sections 4.1.1, 4.2.1). I also made certain assumptions about the presence/absence of edges in my DAG, explained further in Section 4, which might affect the conclusions of my analysis.

3. Methods

3.1 Data Pre-processing and Cleaning

I chose Chaturvedi et al. (2021)'s data set for this project because it contained a manageable amount of data and several relevant variables for my question of interest. Originally, this data set contained 19 variables. For my analysis, I eliminated and combined data columns to form 7 final variables as shown in Table 1: age, school time, entertainment time, fitness time, sleep time, the presence of health issues (treatment), and class rating (outcome). I also re-encoded the data numerically. For example, binary variables encoded as *no/yes* were converted to *0/1*; similarly, class ratings were converted from a five-step *very poor - excellent* scale to an equivalent *1 - 5* numerical scale. All rows with missing data or *n/a* values were dropped, leaving 1108 out of 1182 original data points.

3.2 Graph Elicitation

To learn a graph over the 7 variables described above, I performed causal discovery using Tetrad (Scheines et al., 1998). First, I added a knowledge node with three tiers: 1) age,

Table 1: Variable encodings used to modify Chaturvedi et al. (2021)'s data set

New Variable	Type	Original Variable(s)
Age	Continuous	Age of Subject
School Time	Continuous	Time Spent on Online Class + Time Spent on Self-Study
Entertainment Time	Continuous	Time Spent on TV + Time Spent on Social Media
Fitness Time	Continuous	Time Spent on Fitness
Sleep Time	Continuous	Time Spent on Sleep
Health Issues (Treatment)	Binary: 0/1	Health Issue During Lockdown
Class Rating (Outcome)	Discrete: 1-5	Rating of Online Class Experience

2) entertainment time – fitness time – school time – sleep time, and 3) health issues – class rating. Then, I used the Fast Greedy Equivalence Search (FGES) algorithm (Ramsey et al., 2017), which is an implementation of Greedy Equivalence Search (GES), a score-based method for learning DAGs (Chickering, 2002). FGES was used to learn an equivalence class of possible causal structures with no unmeasured confounders. The Conditional Gaussian Bayesian Information Criterion (CG-BIC) score was selected for model scoring, which relies on parametric assumptions (Haughton, 1988; Schwarz, 1978). I left all the parameters at their default settings but assumed faithfulness and did not discretize continuous variables with discrete children. For bootstrapping, I set the number of re-sampling iterations at 20 and used the preserved ensemble method.

I augmented the resultant DAG from Tetrad by adding, reversing, or removing certain edges using my own background knowledge. A non-parametric conditional independence test known as the Fast Conditional Independence Test (FCIT) (Chalupka et al., 2018) was used to verify that conditional independences implied by the absence of key edges in my graph held in the data being analyzed. FCIT tests the null hypothesis $X \perp\!\!\!\perp Y \mid Z$ and outputs a p-value. If this p-value is less than some significance level α, we reject the null, meaning that the edge $X \rightarrow Y$ should be present. A random seed of 1000 and significance level of $\alpha = 0.05$ was used for all FCIT runs.

3.3 Causal Identification

Next, I used the backdoor criterion, a graphical criterion that adjusts for confounding when data does not come from a perfect randomized controlled trial (Pearl, 1995), to identify the causal effect of health issues on class rating. Essentially, the criterion involves finding a set of variables Z that are sufficient to condition on such that all spurious pathways between the treatment and outcome are blocked.

In particular, I used the SWIG backdoor criterion for my analysis (Richardson and Robins, 2013). This version of the backdoor criterion is identical to the original proposed by Pearl (1995), but in this case, given a DAG $\mathcal{G}$, a SWIG $\mathcal{G}(a)$ is first constructed. Then, a set of variables Z is said to satisfy the backdoor criterion with respect to treatment A and outcome Y in $\mathcal{G}$ if A and $Y(a)$ are d-separated in $\mathcal{G}(a)$; in other words, $A \perp\!\!\!\perp Y(a) \mid Z$ in $\mathcal{G}$.

All variables in Z must also be factual variables in $\mathcal{G}(a)$, not potential outcomes. After identifying a valid backdoor adjustment set Z in this way, I was able to generate the corresponding identifying functional using the following formula, as defined by Pearl (1995):

$$ACE \equiv \mathbb{E}[Y(a) - Y(a')] = \sum_{Z} p(Z) \times \mathbb{E}[Y \mid A = a, Z] - \sum_{Z} p(Z) \times \mathbb{E}[Y \mid A = a', Z]$$

3.4 Causal Estimation

After performing causal identification, I estimated the average causal effect (ACE) of health issues on class rating with my own implementation of augmented inverse probability weighting (AIPW) (Glynn and Quinn, 2010). The AIPW method combines aspects of 1) the outcome regression model $\mathbb{E}[Y \mid A, Z]$ used in the backdoor formula (Pearl, 1995), and 2) the propensity score model $p(A \mid Z)$ used in traditional inverse probability weighting (IPW), which strives to debias the data via re-weighting (Horvitz and Thompson, 1952). This results in a doubly robust estimate for the ACE, protecting against misspecification of either the outcome regression or propensity score model.

I used the following formula to compute the ACE with AIPW given treatment A, outcome Y, and backdoor adjustment set Z (Glynn and Quinn, 2010):

$$\begin{aligned} ACE \equiv \mathbb{E}[Y(a) - Y(a')] = \mathbb{E}\left[\frac{\mathbb{I}(A = a)}{p(A \mid Z)} \times (Y - \mathbb{E}[Y \mid A, Z]) + \mathbb{E}[Y \mid A = a, Z]\right] \\ - \mathbb{E}\left[\frac{\mathbb{I}(A = a')}{p(A \mid Z)} \times (Y - \mathbb{E}[Y \mid A, Z]) + \mathbb{E}[Y \mid A = a', Z]\right] \end{aligned}$$

where $\mathbb{I}(A = a)$ is the indicator function, returning 1 if $A = a$ and 0 otherwise.

I assumed linearity for both models when computing the ACE. I fit a simple logistic regression model for $p(A \mid Z)$ since my treatment, health issues, is binary, and a multinomial logistic regression model for $\mathbb{E}[Y \mid A, Z]$ since my outcome, class rating, had five discrete classes (i.e., 1 - 5). All models were fit using the statsmodels module in Python (Seabold and Perktold, 2010). To quantify uncertainty for causal effect estimates, I computed 95% confidence intervals with 200 bootstraps. A random seed of 0 was used to ensure reproducibility. Causal estimates with and without trimming were also compiled; this heuristic proposed by Crump et al. (2009) is meant to stabilize IPW estimates of the ACE.

4. Results

4.1 The DAG

The final graph elicited for my analysis of the impact on health issues on class rating is illustrated in Figure 1. Pink edges denote the connections that were added or reversed after performing causal discovery with Tetrad.

For example, I reversed the edge between health issues and class rating in $\mathcal{G}$ because I thought it made more sense that a student's health would impact their online class experience rather than the other way around; the *health issues* $\rightarrow$ *class rating* edge is also better suited to address my research question. Additionally, I removed a couple of edges posited

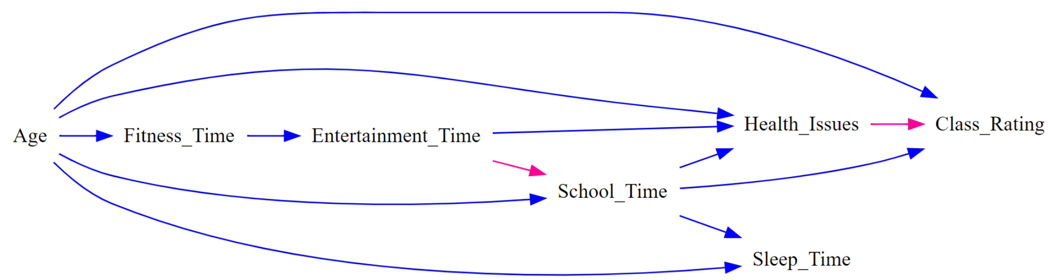

Figure 1: The DAG $\mathcal{G}$ learned over the student survey data by Chaturvedi et al. (2021)

Table 2: FCIT results ($\alpha = 0.05$) for the presence/absence of edges in the learned DAG

Edge	**Z**	**P-value**
Health Issues $\rightarrow$ Class Rating	[Entertainment Time, Age, School Time]	0.0586
Sleep Time $\rightarrow$ Health Issues	[School Time, Age]	0.6584
Fitness Time $\rightarrow$ Class Rating	[Age, Entertainment Time]	0.0949
Entertainment Time $\rightarrow$ Class Rating	[Fitness Time, School Time]	0.9077
Age $\rightarrow$ Class Rating	[Fitness Time, School Time, Health Issues]	0.0004
Entertainment Time $\rightarrow$ School Time	[Fitness Time, Health Issues]	0.0350

by Tetrad, namely the *fitness time* $\rightarrow$ *class rating* and *entertainment time* $\rightarrow$ *class rating* edges, because I did not think these variables would directly impact a student's experience with online learning.

4.1.1 Sensitivity Analysis

FCIT results to confirm the presence/absence of edges in $\mathcal{G}$ are displayed in Table 2. FCIT was performed on the original graph outputted by Tetrad, before the corresponding modifications were made to produce my final DAG (Figure 1). Edges that should remain in $\mathcal{G}$ according to FCIT (i.e., resulted in a p-value ≤ 0.05) are denoted in blue; edges that should be removed from $\mathcal{G}$ (i.e., resulted in a p-value > 0.05) are denoted in red.

FCIT did not agree with all the edges Tetrad initially posited. One reason for this could be that FCIT operates under non-parametric assumptions whereas the FGES causal discovery algorithm I used in Tetrad operates under parametric assumptions (i.e., because of CG-BIC scoring). However, FCIT did confirm that the key *health issues* $\rightarrow$ *class rating* edge should be present, although the p-value (0.0586) was on the borderline, and that edges such as *fitness time* $\rightarrow$ *class rating* and *entertainment time* $\rightarrow$ *class rating* should be deleted. I originally wanted to add edges like *sleep time* $\rightarrow$ *health issues* because it seemed logical that a student's sleep time might impact their health, but since neither FGES nor FCIT supported the presence of this edge, I left it out of my DAG. There was initially an undirected edge between entertainment time and school time as well, but using my FCIT results, I ultimately added the *entertainment time* $\rightarrow$ *school time* edge to $\mathcal{G}$ (Figure 1).

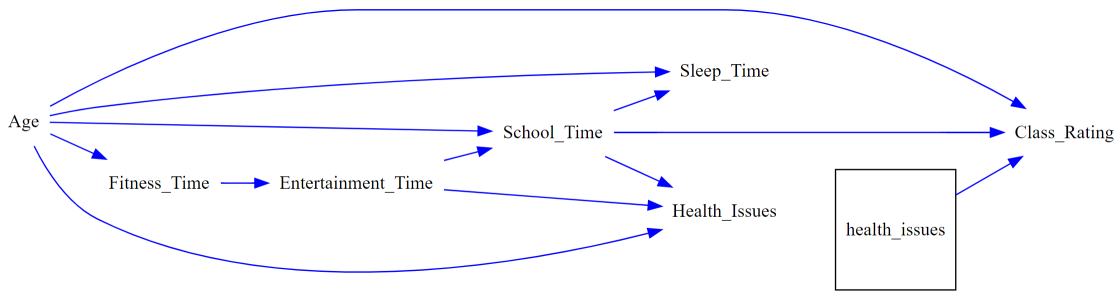

Figure 2: The SWIG $\mathcal{G}$(health issues) corresponding to Figure 1

4.2 Causal Identification and Estimation

Based on my DAG $\mathcal{G}$ (Figure 1), I derived the following backdoor adjustment set $Z = \{\text{Age, School Time}\}$. This is a valid backdoor set because we have that:

$$\text{Health Issues} \perp\!\!\!\perp \text{Class Rating(health issues)} \mid \{\text{Age, School Time}\}$$

in the corresponding SWIG $\mathcal{G}$(health issues) shown in Figure 2. Z is also a minimal and optimal adjustment set. Given Z, we have the following functional for identifying the ACE:

$$\begin{aligned} ACE &\equiv \mathbb{E}[\text{Rating}(1) - \text{Rating}(0)] \\ &= \sum_{\text{Age, School}} p(\text{Age, School}) \times \mathbb{E}[\text{Rating} \mid \text{Health} = 1, \text{Age, School}] \\ &- \sum_{\text{Age, School}} p(\text{Age, School}) \times \mathbb{E}[\text{Rating} \mid \text{Health} = 0, \text{Age, School}] \end{aligned}$$

where variable names have been abbreviated for fit. $Health = 1$ denotes students who experienced health issues during the pandemic; $Health = 0$ denotes students who did not.

This functional was used to estimate the ACE using AIPW. ACE point estimates and 95% confidence intervals are illustrated in Figure 3. AIPW estimates are denoted by orange. Using AIPW without trimming, I obtained an ACE point estimate of -0.5553 and a 95% confidence interval of (-0.7832, -0.3314).

4.2.1 Sensitivity Analysis

I compared the sensitivity of my ACE estimates by first trying various estimator strategies with and without trimming, as depicted in Figure 3. The backdoor, IPW, and AIPW estimators were all implemented using linear models (Section 3.4).

I also tried replacing these linear models with random forest classifiers to compute the ACE with backdoor adjustment and AIPW under alternate assumptions. These results are shown in Table 3, with each point estimate followed by the 95% confidence intervals in parentheses. I used the Random Forest Classifier from scikit-learn library in Python (Pedregosa et al., 2011) to implement these functions. While the ACE estimates varied in magnitude, they were always negative. None of the confidence intervals computed with linear models contained 0 (Figure 3), but the random forest estimates did (Table 3).

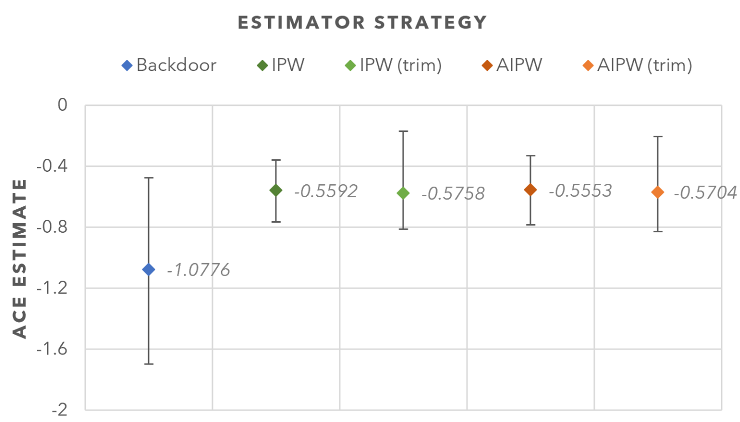

Figure 3: ACE estimates by estimator strategy

Table 3: ACE estimates by linear v.s. random forest models

Estimator Strategy	Model	ACE Estimates
Backdoor	Multinomial Logistic Regression	-1.0776 (-1.6968, -0.4761)
	Random Forest Classifier	-0.2978 (-0.4829, 0.0587)
AIPW	Simple/Multinomial Logistic Regression	-0.5553 (-0.7832, -0.3314)
	Random Forest Classifier	-0.2762 (-0.4788, 0.0797)

5. Discussion

The results in Section 4 demonstrate that the presence of health issues may negatively impact students' online learning experiences. This is supported by the negative ACE estimate of -0.5553 I obtained with untrimmed AIPW (Figure 3). We can interpret this to mean that on average, students experiencing health issues would be expected to rate their online class experience about 0.56 points lower (on a 1-5 scale) than if they did not have health issues. The fact that the AIPW 95% confidence intervals did not contain 0, the null effect value, offers additional evidence in favor of a causal relationship between health issues and class rating. All the other causal estimation strategies I tried with linear models also yielded negative ACE estimates and confidence intervals that did not contain 0 (Figure 3).

Some of these findings may still be sensitive to the faithfulness, linearity, and edge assumptions described in Section 2.1, warranting further investigation. Nonetheless, my sensitivity analysis results from applying different statistical models illustrate that even under non-linear assumptions, a negative causal effect can be obtained. The random forest classifiers did predict a smaller effect than the linear models and produce 95% confidence intervals containing 0 (Table 3), suggesting that perhaps when linearity assumptions are relaxed, there may not be a significant difference between the class ratings of students with and without health issues. I also observed, however, that the ACE estimates from the random forest estimators had higher variances compared with the linear models.

In terms of edge assumptions, FCIT validated the edges I modified from the original graph outputted by Tetrad using my background knowledge (Table 2). The discrepancies between FGES and FCIT illustrate how different causal DAGs can be elicited depending on whether non-parametric or parametric relationships are assumed between variables. While these single edge modifications suggested by FCIT altered the underlying graph, they did not impact the feasibility of identifying and estimating the ACE; they did, however, change the backdoor adjustment set I used in Section 4.2. The presence of key edge *health issues* $\rightarrow$ *class rating* (Figure 1) was also verified by my FCIT results, further supporting the possibility of a causal relationship between these variables.

5.1 Future Work

During my analysis, I considered the possibility of adding bidirected edges to my graph (Figure 1) to account for unmeasured confounding, but in most cases, these edges did not seem to impact the viability of causal identification/estimation. There are a few interesting cases worth exploring though, including the possibility of a *health issues* $\leftrightarrow$ *class rating* edge or adding both the *health issues* $\leftrightarrow$ *sleep time* and *sleep time* $\leftrightarrow$ *class rating* edges. For example, adding the *health issues* $\leftrightarrow$ *class rating* edge would make testing the presence/absence of the *sleep time* $\rightarrow$ *class rating* edge impossible (Figure 1).

I also deleted all rows in Chaturvedi et al. (2021)'s data set with missing data as discussed in Section 3.1, but a more robust analysis might involve augmenting the DAG to reflect this missingness. Future work could involve computing additional causal effects with Chaturvedi et al. (2021)'s data as well. For example, since the original edge posited by Tetrad was *health issues* $\leftarrow$ *class rating* instead of *health issues* $\rightarrow$ *class rating*, it would be interesting to perform a causal analysis in this alternate graph. After all, it is plausible that poor online class experiences could contribute to the presence of student health issues.

Additionally, it would be useful to know which precise health issues impact students' online learning experiences and/or which are the most harmful. It was not clear what exactly the "health issues" variable entailed in Chaturvedi et al. (2021)'s paper (e.g., were these mental or physical health issues?). Having this knowledge would offer greater clarity about how COVID-19 has affected students and help design more effective interventions for improving remote learning experiences.

6. Conclusion

In this project, I performed a causal analysis to investigate how students' health impacts their online learning experiences. My analysis revealed a negative ACE, suggesting that students with health issues may have more negative experiences with remote education. However, these findings may be constrained by the assumptions of faithfulness, linearity, and edge presences/absences; additional analyses should be conducted to verify my results. Future work could also involve augmenting the DAG to account for unmeasured confounders, exploring other causal effects in the data, and investigating more precisely how different health issues impact students' online class experiences.

References

Krzysztof Chalupka, Pietro Perona, and Frederick Eberhardt. Fast conditional independence test for vector variables with large sample sizes. *arXiv preprint arXiv:1804.02747*, 2018.

Kunal Chaturvedi, Dinesh Kumar Vishwakarma, and Nidhi Singh. Covid-19 and its impact on education, social life and mental health of students: A survey. *Children and youth services review*, 121:105866, 2021.

David Maxwell Chickering. Optimal structure identification with greedy search. *Journal of Machine Learning Research*, 3:507–554, 2002.

Richard K. Crump, V. Joseph Hotz, Guido W. Imbens, and Oscar A. Mitnik. Dealing with limited overlap in estimation of average treatment effects. *Biometrika*, 96(1):187–199, 2009.

Adam N Glynn and Kevin M Quinn. An introduction to the augmented inverse propensity weighted estimator. *Political analysis*, 18(1):36–56, 2010.

Dominique MA Haughton. On the choice of a model to fit data from an exponential family. *The annals of statistics*, pages 342–355, 1988.

Daniel G Horvitz and Donovan J Thompson. A generalization of sampling without replacement from a finite universe. *Journal of the American statistical Association*, 47(260): 663–685, 1952.

Judea Pearl. Causal diagrams for empirical research. *Biometrika*, 82(4):669–688, 1995.

Judea Pearl. *Causality*. Cambridge university press, 2009.

Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. Scikit-learn: Machine learning in python. *the Journal of machine Learning research*, 12:2825–2830, 2011.

Joseph Ramsey, Madelyn Glymour, Ruben Sanchez-Romero, and Clark Glymour. A million variables and more: the fast greedy equivalence search algorithm for learning high-dimensional graphical causal models, with an application to functional magnetic resonance images. *International Journal of Data Science and Analytics*, 3(2):121–129, 2017.

Thomas S Richardson and James M Robins. Single world intervention graphs (swigs): A unification of the counterfactual and graphical approaches to causality. *Center for the Statistics and the Social Sciences, University of Washington Series. Working Paper*, 128 (30):2013, 2013.

Richard Scheines, Peter Spirtes, Clark Glymour, Christopher Meek, and Thomas Richardson. The tetrad project: Constraint based aids to causal model specification. *Multivariate Behavioral Research*, 33(1):65–117, 1998.

Gideon Schwarz. Estimating the dimension of a model. *The annals of statistics*, pages 461–464, 1978.

Skipper Seabold and Josef Perktold. Statsmodels: Econometric and statistical modeling with python. In *Proceedings of the 9th Python in Science Conference*, volume 57, page 61. Austin, TX, 2010.