# Causal Structure Discovery of COVID-19 in Chile

Paulo Quinsacara
*Facultad de Ingeniería*
*Universidad Andres Bello*
Santiago, Chile
p.quinsacarajofr@uandresbello.edu

Efraín Campusano
*Facultad de Ingeniería*
*Universidad Andres Bello*
Santiago, Chile
e.campusanomelendez@uandresbello.edu

Pablo Schwarzenberg
*Facultad de Ingeniería*
*Universidad Andres Bello*
Santiago, Chile
pablo.schwarzenberg@unab.cl

Carla Taramasco
*Facultad de Ingeniería*
*Universidad Andres Bello*
Viña del Mar, Chile
carla.taramasco@unab.cl

Billy Peralta
*Facultad de Ingeniería*
*Universidad Andres Bello*
Santiago, Chile
billy.peralta@unab.cl

Hernán Astudillo
*Facultad de Ingeniería*
*Universidad Andres Bello*
Santiago, Chile
hernan.astudillo@unab.cl

***Abstract*—The COVID-19 pandemic has been one of the most severe public health crises of recent decades, exerting profound impacts on population health, economic activity, and social dynamics. In Chile, systematic epidemiological surveillance conducted by the Ministry of Health (MINSAL) during 2020–2021 yielded an anonymized dataset of 54 variables and 6,764,619 patient records. This article reports a study to infer an interpretable causal structure relating demographic, clinical, and health system-related variables, with COVID-19 outcomes, using observational surveillance data. We learn a Directed Acyclic Graph (DAG) with Greedy Equivalence Search (GES), and evaluate its plausibility using a permutation-based falsification test in the DoWhy library, which quantifies violations of the Local Markov Condition (LMC) relative to randomized baselines. The learned GES graph is informative under falsification (0/20 permutations in the Markov equivalence class; p-value = 0.00) and exhibits fewer LMC violations than permuted graphs (79/1629), indicating that it encodes characteristic conditional independencies unlikely to arise by chance. Overall, the resulting structure provides a compact and interpretable causal hypothesis space to support domain interpretation and to prioritize variables and pathways for subsequent causal effect analyses.**

***Index Terms*—Covid 19, causal inference, causal intervention, causal structure, public health.**

## I. Introduction

The COVID-19 pandemic has been one of the most severe public health crises of recent decades, producing substantial health, economic, and social disruption worldwide [1]. In this context, scientific evidence has been essential for understanding the multifactorial determinants of pandemic outcomes and for supporting decision-making under uncertainty.

In Chile, pronounced socioeconomic stratification and territorial heterogeneity were reflected in markedly unequal patterns of COVID-19 incidence and mortality across population groups and municipalities [2]. To monitor these dynamics, Ministry of Health (MINSAL) implemented large-scale epidemiological surveillance during 2020 to 2021, supported by national reporting systems such as EPIVIGILA, which served as a primary source for official reports and high-granularity case-level data [3], [4]. This surveillance provides a valuable empirical basis to study clinical trajectories and outcomes including hospitalizations, intensive care admissions, and mortality across successive pandemic waves.

A central challenge is to move from association to causation. While descriptive and correlational analyses are informative, they do not, by themselves, distinguish spurious associations from causal mechanisms. This distinction is critical for public policy because decisions grounded on non-causal evidence can lead to ineffective or counterproductive interventions. Motivated by this gap, this work focuses on causal structure discovery from observational surveillance data. We learn a Directed Acyclic Graph (DAG) using Greedy Equivalence Search (GES), a score-based method that searches over equivalence classes of causal graphs [5]. To assess whether the learned structure is meaningfully consistent with the observed conditional independencies, we apply a permutation-based falsification test implemented in DoWhy[1] [6]. This test evaluates Local Markov Condition violations and compares them against randomized baselines, yielding an interpretable measure of whether the proposed DAG performs better than chance [7].

The remainder of this article is structured as follows: Section II surveys related work; Section III presents causal graphical models, GES, and permutation-based falsification; Section IV describes dataset, preprocessing, and experimental setup; Section V discusses the learned causal structure in the COVID-19 setting; and Section VI summarizes and concludes.

## II. Related work

Causal reasoning has been widely adopted during the COVID-19 pandemic to move beyond correlation and support policy-relevant interpretation under observational data constraints. In the COVID-19 literature, a common approach is to posit a causal graph (or a set of adjustment assumptions) and then estimate the effects of non-pharmaceutical interventions, mobility restrictions, or testing strategies. For instance, policy-oriented causal analyses have been conducted

[1]DoWhy: https://github.com/py-why/dowhy

for Chile, emphasizing heterogeneous effects across territories and socioeconomic contexts [8], [9]. Complementarily, causal graph analyses have also been used to study drivers of reported case numbers in other national settings, illustrating how DAG-based reasoning can structure the analysis of observational epidemiological data [10]. These studies motivate the need for principled causal modeling in pandemic surveillance, but most of them rely on pre-specified structures or targeted effect estimation rather than learning an interpretable causal structure directly from high-dimensional surveillance variables.

Causal structure learning (or causal discovery) in health and biomedicine has grown rapidly, spanning constraint-based, score-based, and more recent machine learning approaches. A scoping review by Upadhyaya et al. synthesizes traditional and deep-learning causal discovery methods and discusses practical issues that are especially salient in biomedical settings, including scalability, mixed data types, measurement noise, and the mismatch between algorithmic assumptions and real-world data generation [11]. Within electronic health record (EHR) contexts, causal discovery has been explored to support clinical decision-making; for example, Shen et al. propose and evaluate a causal structure discovery workflow on EHR data for clinical use cases, explicitly contrasting design choices and practical constraints found in healthcare data [12]. These works underline the promise of learning causal hypotheses from observational health records, but also highlight persistent challenges such as latent confounding, selection bias, and interpretability for domain stakeholders.

Within the family of structure learning approaches, score-based search over Markov equivalence classes is often attractive for interpretability and computational reasons. GES (Greedy Equivalence Search) is a canonical score-based algorithm that aims to identify an equivalence class of DAGs maximizing a chosen score under standard assumptions (e.g., causal sufficiency and faithfulness). However, real-world health data frequently violate idealized assumptions, which can yield unstable or non-identifiable structures. Recent work emphasizes open methodological problems in causal structure learning and stresses that careful validation, robustness checks, and domain constraints are central when translating causal discovery outputs into actionable scientific hypotheses [13]. Relatedly, incorporating prior knowledge has been proposed as a way to restrict the search space and improve plausibility in applied healthcare causal discovery, as illustrated by frameworks that combine domain constraints with discovery algorithms [14].

A key gap in applied causal discovery is how to evaluate whether a learned graph is meaningfully compatible with the observed data when ground-truth causal graphs are unavailable. A promising line of work proposes falsification-style checks that compare a candidate DAG against randomized baselines. In particular, permutation-based falsification tests operationalize compatibility through violations of the Local Markov Condition and quantify whether the candidate graph encodes conditional independencies that are unlikely to arise by chance [15]. DoWhy provides an end-to-end causal inference framework that includes refutation and falsification utilities that facilitate such checks in practice [6].

## III. Background

Our work focuses on learning a compact and interpretable causal hypothesis space from Chilean COVID-19 surveillance variables by using GES, and assessing plausibility via permutation-based falsification testing. This design choice explicitly prioritizes interpretability and empirical falsifiability of the learned structure as prerequisite for subsequent domain interpretation and downstream causal effect analyses.

### A. Causal models

Causal modeling extends classical statistical analysis by representing not only associations among variables but also hypothesized data generating mechanisms that support cause and effect interpretation. In contrast to descriptive approaches, which typically capture symmetric relationships, causal models encode asymmetric dependencies and enable reasoning about interventions, counterfactual questions, and the propagation of changes across a system [16].

A central representation in causal inference is the directed acyclic graph (DAG), where nodes correspond to variables and directed edges denote direct causal influences. Formally, a graph is defined as $G = (V, E)$, with a set of nodes $V$ and directed edges $E$. In causal structure learning and causal graphical models, the terms "node" and "variable" are commonly used interchangeably [17].

Under the causal Markov and faithfulness assumptions, graphical separation in a DAG corresponds to conditional independence in the observed distribution [16]. This correspondence is characterized by d-separation. Intuitively, two variables $X$ and $Y$ are d-separated by a conditioning set $Z$ if every path between $X$ and $Y$ is blocked given $Z$; otherwise they are d-connected.

A path is blocked given $Z$ according to these rules [16]:

1) **Chains and forks are blocked by conditioning:** for a chain $X \to M \to Y$ or a fork $X \leftarrow M \to Y$, conditioning on the middle node $M \in Z$ blocks the path; if $M \notin Z$, the path remains potentially active.
2) **Colliders are blocked unless conditioned on, or conditioned on descendants:** for a collider $X \to M \leftarrow Y$, the path is blocked if neither $M$ nor any descendant of $M$ is in $Z$. Conditioning on $M$, or on a descendant of $M$, opens the path and can induce dependence, often referred to as collider bias.

In the causal interpretation of a DAG $G$, directed edges represent direct causal relations, while d-separation provides the link between graph structure and conditional independencies used for discovery, validation, and adjustment reasoning.

### B. Causal structure discovery algorithms

We perform causal structure discovery using the Greedy Equivalence Search (GES) algorithm [5], a score-based algorithm that searches over Markov equivalence classes of directed acyclic graphs (DAGs). Given a scoring criterion, such as the Bayesian Information Criterion (BIC), GES aims

to identify a graph that optimizes the score under standard assumptions used in causal discovery, including causal sufficiency and faithfulness. GES has two phases:

- **Forward**: starting from an empty graph, edges are added greedily. At each step, the candidate edge addition that yields the largest improvement in the BIC score is selected, and the process continues until no addition improves the score.
- **Backward**: starting from the graph obtained in the forward phase, edges are removed greedily. At each step, the candidate edge removal that yields the largest improvement in the BIC score is selected; the process stops when no removal improves the score.

The output is a compact representation of the dependency structure implied by the data. Because the result is defined up to Markov equivalence, some edge orientations may remain non-identifiable from observational data alone; nevertheless, the learned structure provides an interpretable causal hypothesis space for downstream analysis and domain interpretation.

### C. Falsification testing based on permutations

To assess whether the learned DAG is compatible with the conditional independencies present in the data, we use the falsification procedure provided by DoWhy [6]. It is based on the Local Markov Condition (LMC), which states that each node is conditionally independent of its non descendants given its parents in the DAG [16]. Operationally, the method runs conditional independence tests implied by the candidate DAG, and computes a violation score for it by counting how many implications are violated in the observed data [7].

To make this score interpretable, we construct a null distribution via permutations. Specifically, we generate $B = 20$ random permutations of the node labels, keeping the graph pattern fixed, and rerun the same LMC based testing procedure on each permuted DAG [7]. For each permutation, we record the total number $v_{\text{obs}}$ of LMC violations, and compare it against the permutation distribution. DoWhy reports a permutation p value that quantifies how often a randomized relabeling yields a violation count at least as good as the observed DAG.

In our experiment, the learned GES DAG is reported as informative because 0 out of 20 permutations fall within the Markov equivalence class of the given DAG, with a reported p value of 0.00. In addition, the learned DAG violates 79 out of 1629 tested LMC implications and performs better than 100% of the permuted DAGs, with a reported $p = 0.00$. Based on the chosen significance level of 0.05 and the informativeness result, we do not reject the learned DAG.

This falsification analysis should be interpreted as an empirical compatibility check between the proposed DAG and the observed conditional independencies. It does not prove causal correctness and it does not rule out latent confounding, measurement error, or selection effects.

### D. Assumptions and limitations

Our causal structure learning analysis relies on standard assumptions commonly used in observational causal discovery. First, causal sufficiency is assumed, meaning that the main common causes of the measured variables are observed in the dataset; unmeasured confounding can bias both adjacency and edge orientation. Second, the causal Markov and faithfulness assumptions are adopted to relate graphical separation to conditional independence relations in the data. Third, the learned structure is defined up to Markov equivalence, so some edge orientations may be non-identifiable from observational data alone. Finally, because the variables come from surveillance and clinical reporting workflows, measurement and selection effects may induce dependencies that reflect recording mechanisms rather than biological causation. For these reasons, the learned DAG should be interpreted as a compact causal hypothesis space supported by the observed conditional independencies, rather than a definitive causal ground truth.

## IV. Experiments

We conducted several experiments.

### A. Dataset and preprocessing

The dataset used in this study contains 54 variables and 6,764,619 patient level records collected through national epidemiological surveillance in Chile during 2020–2021. Prior to analysis, records with incomplete information were removed to ensure consistent input for score based structure learning and falsification testing. The final set of variables used in the analysis is listed in Table I.

### B. Causal structure learning and falsification setup

We learn a directed acyclic graph (DAG) using Greedy Equivalence Search (GES) and score the candidate structures using a BIC based criterion. All experiments were implemented in Python, and the permutation based falsification analysis was conducted using the DoWhy library[2]. For falsification, we evaluate Local Markov Condition implications and compare the observed number of violations against a null distribution obtained from $B = 20$ random permutations of node labels.

### C. Application of algorithms for structure causal discovery

Fig. 1 shows the directed acyclic graph (DAG) learned with the GES algorithm from the MINSAL surveillance dataset. The resulting structure is compact and interpretable, providing a structured view of dependencies among demographic, clinical presentation, diagnostic, and outcome-related variables. In particular, the graph organizes symptom variables in relation to diagnostic and status indicators, which facilitates domain interpretation and communication of plausible pathways to decision-makers based on observed conditional dependencies in the data.

[2]https://github.com/py-why/dowhy

TABLE I: Variables and their descriptions.

| Variable | Description | Variable | Description |
|---|---|---|---|
| EPIDEMIOLOGICAL-WEEK | Epidemiological week of case notification | CLINICAL-STAGE | Clinical stage of the patient |
| REGION | Administrative region code | PATIENT-STATUS | Current status of the patient |
| UNDERLYING-CAUSE-OF-DEATH | Registered underlying cause of death | SEX | Biological sex of the patient |
| AGE | Age of the patient in years | PREGNANCY | Pregnancy status of the patient |
| HEALTH-INSURANCE | Health insurance system affiliation | NATIONALITY | Nationality of the patient |
| REGION-OF-RESIDENCE | Region where the patient resides | CLINICAL-PRESENTATION | Initial clinical presentation of symptoms |
| EPIDEMIOLOGICAL-WEEK-FIRST-SYMPTOMS | Epidemiological week of symptom onset | ALCOHOL-CONSUMPTION | Alcohol consumption status |
| TOBACCO-USE | Tobacco consumption status | VAPING-USE | Use of electronic cigarettes or vaping devices |
| INTERNATIONAL-TRAVEL-HISTORY | History of international travel prior to diagnosis | CONTACT-WITH-SUSPECTED-CASE | Contact with a suspected COVID-19 case |
| ANTIPYRETIC-MEDICATION-USE | Use of antipyretic medication | HOSPITALIZATION | Hospitalization status of the patient |
| CONTACT-WITH-SICK-PEOPLE | Contact with individuals presenting symptoms | HEALTHCARE-WORKER | Healthcare worker occupational status |
| ICU-OR-INTERMEDIATE-CARE-ADMISSION | Admission to intensive or intermediate care unit | DIAGNOSIS | Medical diagnosis status |
| LABORATORY-CONFIRMATION | Laboratory confirmation of COVID-19 | CLINICAL-CONFIRMATION | Clinical confirmation of COVID-19 |
| CASE-STATUS | Epidemiological case status | CLOSED-CASE | Case closure status |
| CASE-DETECTION-TYPE | Case detection or surveillance type | PLACE-OF-REST | Location where the patient rested during illness |
| RISK-FACTOR-CONFIRMED-CASE | Presence of risk factors for confirmed cases | RISK-FACTOR-SUSPECTED-CASE | Presence of risk factors for suspected cases |
| INTERNATIONAL-TRAVEL-RISK-FACTOR | Risk factor related to international travel | NATIONAL-TRAVEL-RISK-FACTOR | Risk factor related to domestic travel |
| NO-RISK-FACTOR | Absence of identified risk factors | ORD-STATUS | Ordinal health status classification |
| DECEASED | Mortality outcome indicator | CURRENT-CASES | Indicator of currently active cases |
| EPIVIGILA-PCR | PCR test result recorded in Epivigila | HEADACHE | Presence of headache symptom |
| MYALGIA | Presence of muscle pain (myalgia) | DYSPNEA-OR-RESPIRATORY-DIFFICULTY | Presence of dyspnea or breathing difficulty |
| LOSS-OF-SMELL-ANOSMIA | Presence of loss of smell (anosmia) | LOSS-OF-TASTE-AGEUSIA | Presence of loss of taste (ageusia) |
| COUGH | Presence of cough symptom | FEVER | Presence of fever symptom |
| CHEST-PAIN | Presence of chest pain | OTHER | Presence of other reported symptoms |
| SORE-THROAT | Presence of sore throat (odynophagia) | ABDOMINAL-PAIN | Presence of abdominal pain |
| PROSTRATION | Presence of severe fatigue or prostration | DIARRHEA | Presence of diarrhea |
| TACHYPNEA | Presence of abnormally rapid breathing | CYANOSIS | Presence of cyanosis |


Fig. 1: Causal graph generated by the GES algorithm. For readability, the learned DAG is presented on a dedicated landscape page at high resolution.

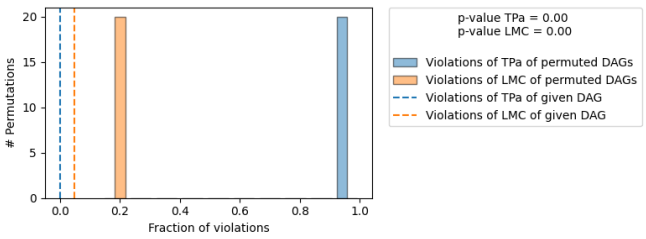


Fig. 2: Result of the falsification test using the GES algorithm.

### D. Assessing DAG Validity through Falsification Testing

To rigorously assess the validity of a specified DAG, we utilize DoWhy's falsification framework. The core of this procedure is to evaluate violations of the Local Markov Condition (LMC), that states that a node it is independent of other nodes in the DAG, except their descendants, given their parent nodes. This condition is assessed by performing conditional independence (CI) tests for each node given its parents in the candidate graph [7].

Fig. 2 shows the GES algorithm results from the falsification test after its application: the provided DAG is informative since 0 / 20 of the permutations lie in the Markov equivalence class of the given DAG (p-value: 0.00); and he provided DAG violates 79/1629 LMCs and performs better than 100.0 percent of the permuted DAGs (p-value: 0.00). Based on the provided significance level (0.05), and because the DAG is informative, we do not reject the DAG.

### E. Robustness considerations

As an empirical robustness check in the absence of ground truth, we rely on permutation-based falsification to evaluate whether the learned structure is more compatible with the observed conditional independencies than randomized baselines. Additional robustness analyses such as edge stability under resampling, sensitivity to scoring choices, and comparisons against alternative discovery families are left for future work.

## V. COVID-19 Interpretation of Causal Structure

The obtained causal structure enables interpretation.

### A. Interpretation of the GES estimated causal graph

The causal graph was estimated using Greedy Equivalence Search (GES), which identifies a directed acyclic graph that optimizes a penalized goodness of fit criterion, typically BIC, under acyclicity and standard structural assumptions. Because observational structure learning generally identifies a Markov equivalence class, the directed edges should be interpreted as data supported direct dependency hypotheses, while recognizing that some orientations can be non identifiable without additional domain constraints or temporal information.

Overall, the learned DAG exhibits a structured organization rather than an arbitrary dense network. Three coherent subgraphs are apparent: (1) a **clinical symptom block**, grouping respiratory and systemic manifestations and connects to higher level clinical coding variables, indicating consistent conditional dependency patterns among symptoms and clinical presentation; (2) a **surveillance and confirmation core**, linking diagnosis, PCR-related variables, laboratory and clinical confirmation, and case status and closure, suggesting that a substantial portion of the learned dependencies reflects the operational workflow of case detection and classification in surveillance data; and (3) a **severity and outcome axis**, connecting hospitalization and intensive or intermediate care admission with status and mortality related variables. Notably, the graph includes a directed link from deceased to patient status, consistent with status encoding in administrative records.

### B. Temporal and operational structure in surveillance variables

Among structural variables, *epidemiological week* has high connectivity and links to several clinical and administrative nodes. This pattern should not be interpreted as a biological causal factor, since it is consistent with a temporal system level proxy that captures non-observed changes across the pandemic (e.g. shifts in testing capacity, diagnostic criteria, reporting practices, and health system pressure). Thus, *epidemiological week* acts as a compact summary of time varying mechanisms that shape the joint distribution of recorded variables.

Two variables emerge as high connectivity hubs: *case detection type* and *clinical presentation*. They are bridges between symptom information and the confirmation and classification subsystem, supporting the interpretation that detection and recording mechanisms are central drivers of the dependency structure captured from observational surveillance variables.

### C. Clinical progression and severe outcomes

The severity and outcome subgraph provides a clinically plausible dependency pathway linking clinical coding variables with hospitalization and intensive or intermediate care admission, and then with status and mortality-related variables. This structure is consistent with the real world flow of COVID-19 case management, where increasing clinical severity is associated with higher probability of hospitalization and care escalation. In the learned graph, outcome related variables tend to appear downstream of multiple clinical and administrative factors, which supports interpreting them as terminal endpoints integrating accumulated influences along the recorded care trajectory.

### D. Demographic and administrative dependencies

The graph also contains a demographic and administrative chain from region of residence to sex to pregnancy to nationality to health insurance, highlighting strong dependencies among background variables. At the same time, such orientations in exogenous attributes should be interpreted cautiously, because they can arise from Markov equivalence and the observational nature of the data rather than literal biological causation. Age also appears as a structurally important variable with connections to multiple severe clinical and outcome related nodes, consistent with its role as a broad confounder

that jointly influences several stages of disease progression and clinical decision making.

### E. Laboratory confirmation and potential testing bias

Laboratory confirmation appears as a mediator between clinical presentation, diagnosis, and downstream clinical decisions, suggesting that confirmation is not an exogenous measurement step in the recorded process. This pattern is compatible with severity conditioned testing, where more severe cases have higher probability of being tested and confirmed, which can shape the observed dependency structure in surveillance data.

### F. Global assessment supported by falsification

Permutation based falsification supports the informativeness of the learned structure. Relative to permuted baselines, the observed DAG achieves near zero violations under TPa and a low fraction of Local Markov Condition violations, with reported p values equal to 0.00 for both metrics. Therefore, the learned GES graph encodes conditional independencies unlikely to arise by chance, providing a compact and interpretable causal hypothesis space for subsequent domain grounded analyses, while not constituting proof of causal correctness under unmeasured confounding or selection effects.

## VI. Conclusions

This work demonstrates the practical value of discovering the causal structure for a large-scale public health surveillance problem. Using Chilean COVID-19 case-level surveillance data from 2020–2021 (54 variables; 6,764,619 records), we learned an interpretable directed acyclic graph (DAG) with Greedy Equivalence Search (GES) and evaluated its plausibility via permutation-based falsification in DoWhy. The learned DAG was informative under falsification (0/20 permutations in the Markov equivalence class; reported p-value = 0.00) and exhibited substantially fewer Local Markov Condition violations than randomized baselines (79/1629, reported p-value = 0.00), supporting that the structure encodes conditional independencies unlikely to arise by chance.

Clearly, the learned graph highlights a clear structural pattern around the role of diagnosis within the COVID-19 clinical and surveillance system, particularly with respect to severe outcomes. Diagnosis does not appear as an isolated exogenous factor, but rather as an intermediate node embedded in a broader recorded trajectory shaped by clinical stage, age, and time varying system factors captured by epidemiological week. In the learned structure, diagnosis is positioned together with critical downstream care variables such as hospitalization and intensive or intermediate care admission, which in turn connect to terminal outcome variables related to patient status and mortality. This configuration is consistent with interpreting diagnosis as a key mediator that links early clinical presentation and coding to clinical decisions that accompany progression toward severe outcomes.

At the same time, the graph suggests that diagnosis is itself influenced by clinical severity and structural temporal factors, indicating nontrivial dependence introduced by healthcare system dynamics and reporting practices. Therefore, associations between diagnosis and mortality should not be interpreted as a direct causal effect. Instead, they are better understood as part of a sequential mechanism, where diagnosis triggers or enables subsequent interventions and care pathways that may modify the risk of adverse outcomes. In general, mortality emerges as a terminal endpoint integrating multiple pathways, reinforcing that the fatal outcomes of COVID-19 in the surveillance data are multicausal and shaped by demographic factors, clinical progression, and previous diagnostic and care decisions.

From a practical perspective, the learned structure helps distinguish clinical pathways from surveillance and recording mechanisms, which is relevant for policy interpretation. In particular, the central role of diagnosis and confirmation variables suggests that changes in testing capacity, diagnostic criteria, and case detection policies can alter observed dependencies with severe outcomes, highlighting the need to interpret surveillance associations in light of operational context. Importantly, falsification provides an empirical compatibility check rather than proof of causal correctness. The learned structure may still be affected by latent confounding, measurement error, selection effects, and non-identifiability within Markov equivalence classes. Nevertheless, the resulting DAG provides a compact and empirically supported causal hypothesis space that can guide domain interpretation and prioritize variables and pathways for subsequent causal effect estimation and sensitivity analyses.

Future work will incorporate additional causal discovery algorithms, including constraint-based and hybrid approaches, to enable systematic comparisons of learned structures and falsification outcomes under the same data and evaluation protocol, and to assess edge stability under resampling.

## References


[1] International Labour Organization, Food and Agriculture Organization of the United Nations, International Fund for Agricultural Development, and World Health Organization, "Impact of covid-19 on people's livelihoods, their health and our food systems," Joint statement, Oct. 2020, accessed: 2026-01-31. [Online]. Available: https://www.who.int/news/item/13-10-2020-impact-of-covid-19-on-people%27s-livelihoods-their-health-and-our-food-systems

[2] G. E. Mena, P. P. Martinez, A. S. Mahmud, P. A. Marquet, C. O. Buckee, and M. Santillana, "Socioeconomic status determines COVID-19 incidence and related mortality in Santiago, Chile," *Science*, vol. 372, no. 6545, p. eabg5298, May 2021.

[3] Ministerio de Salud de Chile (MINSAL), "Epivigila: Sistema nacional integrado de vigilancia epidemiológica (plataforma web)," Online system portal, n.d., accessed: 2026-01-31. [Online]. Available: https://epivigila.minsal.cl/

[4] C. Taramasco, C. Rimassa, J. Acevedo Romo, A. Césped Zavando, and R. Fuentes Bravo, "Vigilancia epidemiológica en pandemia de covid-19: Sistema epivigila," *Medwave*, vol. 22, no. 5, p. 002560, 2022, published: Jun 2, 2022. [Online]. Available: https://www.medwave.cl/revisiones/analisis/8742.html

[5] D. M. Chickering, "Optimal structure identification with greedy search," *Journal of Machine Learning Research*, vol. 3, pp. 507–554, 2002.

[6] A. Sharma and E. Kiciman, "Dowhy: An end-to-end library for causal inference," *arXiv preprint arXiv:2011.04216*, 2020.

[7] E. Eulig, A. A. Mastakouri, P. Blöbaum, M. Hardt, and D. Janzing, "Toward falsifying causal graphs using a permutation-based test," *Machine Learning*, 2023.

[8] M. Bennett, "All things equal? heterogeneity in policy effectiveness against covid-19 spread in chile," *World Development*, vol. 137, p. 105208, 2021.
[9] W. Kristjanpoller, K. Michell, and M. C. Minutolo, "A causal framework to determine the effectiveness of dynamic quarantine policy to mitigate covid-19," *Applied Soft Computing*, vol. 104, p. 107241, 2021.
[10] E. Steiger, T. Mussgnug, and L. E. Kroll, "Causal graph analysis of covid-19 observational data in german districts reveals effects of determining factors on reported case numbers," *PLOS ONE*, vol. 16, no. 5, p. e0237277, 2021.
[11] P. Upadhyaya, K. Zhang, C. Li, X. Jiang, and Y. Kim, "Scalable causal structure learning: Scoping review of traditional and deep learning algorithms and new opportunities in biomedicine," *JMIR Medical Informatics*, vol. 11, p. e38266, 2023.
[12] Y. Shen, S. Ma, P. Vemuri, M. R. Castro, P. J. Caraballo, and G. J. Simon, "A novel method for causal structure discovery from electronic health record data for clinical decision support," *Scientific Reports*, vol. 11, p. 21025, 2021.
[13] A. C. Constantinou, A. Hasan, L. McCracken, and N. E. Fenton, "Open problems in causal structure learning: A case study of covid-19 in the uk," *Expert Systems with Applications*, vol. 234, p. 121069, 2023.
[14] A. Hasan, R. Mbuvha, A. Constantinou, and A. A. Freitas, "Kcrl: A prior knowledge-based causal discovery framework for healthcare data," in *Proceedings of Machine Learning for Health (ML4H)*, ser. Proceedings of Machine Learning Research, vol. 182, 2022.
[15] E. Eulig, A. A. Mastakouri, P. Blöbaum, M. Hardt, and D. Janzing, "Toward falsifying causal graphs using a permutation-based test," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 39, no. 25, pp. 26 778–26 786, 2025.
[16] J. Pearl, *Causality*. Cambridge university press, 2009.
[17] A. M. Lipsky and S. Greenland, "Causal directed acyclic graphs," *JAMA*, vol. 327, no. 11, pp. 1083–1084, 2022.