Causal Model Analysis of Infection Impact on Health Trajectory Using Symptom and Biometric Data

Rachel Robles

Supervisor(s): Dr. Ioannis Papathanail, Prof. Dr. Stavroula Mougiakakou
Institution(s): University of Bern, ARTORG Center for Biomedical Engineering Research

Introduction

As COVID-19 reinfections rise following the relaxation of public health measures, there is growing urgency to understand how repeated infections influence long-term health—especially in populations with chronic illness, such as Long COVID and Myalgic Encephalomyelitis/Chronic Fatigue Syndrome (ME/CFS). Despite evidence of worsening long-term health outcomes following SARS-CoV-2 reinfections, existing research has not investigated the effect of cumulative infection burden or established causal insights. The aim of this thesis was to determine the causal influence of cumulative COVID-19 infections on the quantity, severity, and frequency of symptoms in this chronically ill population.

Materials and Methods

Using symptom, biometric, exertion, and demographic data from over 11,000 Visible app participants who self-reported ~2,500 SARS-CoV-2 infection events, a dataset was constructed representing pre-, acute, and post-infection phases. Five causal discovery algorithms were applied to this data and evaluated at three edge-frequency thresholds, yielding fifteen candidate causal graphs and Structural Causal Models (SCMs). The graphs/SCMs were assessed using node-level statistics, structural stability, and compatibility with algorithmic assumptions to select an optimal model for causal inference.

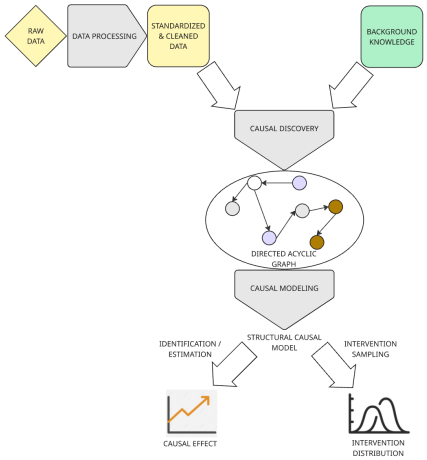

Fig. 1 Overview of the causal modeling pipeline from raw data collection through analysis of causal effects and intervention sampling. The directed acyclic graph represents causal relationships identified through causal discovery algorithms, which are then used to estimate causal effects and simulate intervention distributions.

The SCM simulated interventions on cumulative infections to estimate symptom outcomes. The final model was evaluated for stability, feasibility, and interpretability.

Results

Intervention sampling on a NOTEARS [1] nonlinear graph revealed that increasing infection episodes from 1 to 6 produced increases in post-infection symptom severity and resting heart rate by 4.5% and 5.9%, respectively. As sleep quality declined, post-infection symptom severity increased by 124% and symptom frequency by 51%. Emotional stress influenced post-infection symptom severity and frequency by 48.0% and 23.7%, respectively. Outgoing edges from infection episode demonstrated a reinfection's effects on post-infection symptoms were mediated primarily through cardiovascular dysregulation.

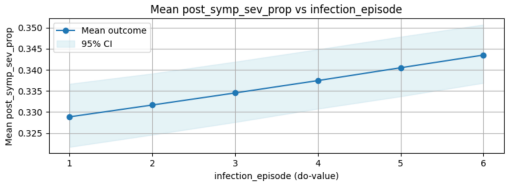

Fig. 2 Visualization of Monte Carlo symptom severity means and confidence intervals for each intervention value of infection episode (from 1 to 6).

Discussion

These results demonstrate that while reinfections exert measurable causal effects on symptom severity and physiology, pre-infection stress, sleep quality, and autonomic function contribute more strongly to post-infection outcomes. This is among the first studies to apply causal discovery to reinfection-specific longitudinal data in chronic illness populations. It offers actionable insights for personalized health management and reinfection risk mitigation, and serves as a blueprint for finding causal associations from digital health data.

References

[1] X. Zheng, B. Aragam, P. Ravikumar, and E. P. Xing, DAGs with NO TEARS:
Continuous Optimization for Structure Learning, arXiv:1803.01422 [stat], Nov. 2018.

Acknowledgements

The project was supported by Patient Led Research Collaborative and Visible Health Inc. Special thanks to Dr. Megan Fitzgerald and Rooholla Poursoleymani for biomedical and technical expertise.