

International Journal of Population Data Science

Journal Website: www.ijpds.org



A causal inference framework for individual-level models.

Lynette Linzbuoy¹, Mark S Gilthorpe², Alison J Heppenstall³, and Jiaqi Ge⁴

¹Leeds Institute for Data Analytics, University of Leeds, Leeds, United Kingdom

²Obesity Institute, Leeds Beckett University, Leeds, United Kingdom

³School of Political and Social Sciences, University of Glasgow, Glasgow, United Kingdom

⁴School of Geography, University of Leeds, Leeds, United Kingdom

Introduction & Background

Most human beings have an intuitive understanding of causation; however, it is a complex phenomenon which remains largely under-researched. After decades of research and philosophical discussion, there are different formal systems that seek to define what the concepts of causality mean and, importantly, how we might understand causality within practical research applications.

To address this, many methods have been developed across different fields. In epidemiology, the main approach relies on the evaluation of counterfactual contrasts via statistical regression models informed by either the theoretical potential outcomes framework or graphical causal models (in the form of a directed acyclic graph, DAG). In recent years, agent-based models (ABMs) have emerged as promising tools for causal inference evaluations within complex systems.

Objectives & Approach

The aim was to build a causal framework embedding a priori causal structures into synthetic data for robust, causally informed ABMs.

Key variables identified in the DAG included age, sex, ethnicity, and comorbidities, which directly influenced infection susceptibility and recovery. Synthetic data were generated based on these predefined causal relationships. These data were embedded within a NetLogo ABM, where agent infection statuses were updated iteratively using logistic regression models grounded in DAG-defined paths. The DAG-informed ABM was compared with a naïve ABM, where transition parameters were selected from established infectious disease models, and a spatially unconstrained microsimulation model (MSM), where transition parameters were derived explicitly from the causally structured synthetic data, but agent-to-agent interaction was not permitted.

Relevance to Digital Footprints

Through the understanding of individual behaviours, a directed acyclic graph is generated to codify important variables influencing infectious disease prevalence. The synthetic data generated from this enables a realistic simulation of population health patterns, thereby identifying key variables for informed public health interventions. These methods could be applied to digital footprint data of various types.

Conclusions & Implications

Integrating DAGs with ABMs enhances the robustness of disease transmission simulations. The DAG-informed ABM produced outcomes more closely aligned with expected epidemiological behaviour, such as heterogeneous infection and recovery patterns, including subpopulations that remained uninfected throughout the simulation. Future work will extend the framework to incorporate real-world spatial data and multilevel hierarchical structures.

Keywords

causal inference, directed acyclic graphs, agent-based modelling, simulation studies

