

---

# How to model COVID-19 pathophysiology during lockdown: remote elicitation of causal Bayesian networks from medical experts

---

Steven Mascaro<sup>1</sup>   Yue Wu<sup>2</sup>   Owen Woodberry<sup>1</sup>   Erik P. Nyberg<sup>1</sup>   Jessica Ramsay<sup>3</sup>   Ross Pearson<sup>1</sup>  
Ann E. Nicholson<sup>1</sup>   Tom Snelling<sup>2</sup>

<sup>1</sup> Department of Data Science & AI, Monash University, Melbourne, Australia

<sup>2</sup> School of Public Health, University of Sydney, Sydney, Australia

<sup>3</sup> Wesfarmers Centre of Vaccines and Infectious Diseases, Telethon Kids Institute, Perth, Australia

## AIM: CAUSAL BNS FOR COVID PHYSIOLOGY

COVID-19 pathophysiology—the way the disease develops internally—is unusual in several respects, including its propensity to cascading complications and relatively high fatality rate. Given its persistent global presence, better understanding of these causal processes remains urgent and important for improving patient outcomes. In early 2020, as the pandemic commenced, we began a major project to model this pathophysiology.

We used Bayesian networks (BNs) as our models, because they provide perspicuous maps of probabilistic relationships and powerful tools for calculation. We used *causal* BNs, because causality underpins diagnosis (what caused the symptoms), prognosis (its likely further effects), and treatment (the likely effect of an intervention). Hence, AI that uses causal information for diagnosis can perform better [Richens et al., 2020], and non-causal models cannot predict the effects of novel interventions at all [Korb and Nicholson, 2011]. Where necessary, e.g., for COVID complications, we used *dynamic* causal BNs, which can include feedback loops by representing successive “time slices” of the system. BNs have proved to be a powerful tool in many domains, including medicine. Dynamic causal BNs have been used to model the epidemiology of COVID [Friston et al., 2020], but our project is the first to use them for clinical purposes.

## METHOD: REMOTE EXPERT ELICITATION

The nature of the disease and the worldwide response placed unusual constraints on our methods. However, our adaptations also proved to have instructive advantages.

**Domain experts filter literature:** Early in the pandemic, extensive publicly available patient datasets did not exist (and even now, remain relatively sparse), which ruled out discovering causal structure through purely automated means. In contrast, the medical literature was flooded with unfiltered, technical and sometimes conflicting preprints. Thus, the only feasible option was to elicit causal structure from domain experts. We used groups of volunteer, independent,

medical specialists to filter, interpret and discuss the literature findings and develop a reasonable current consensus. Furthermore, we utilized intermediary experts with some combined domain and BN expertise to facilitate the exchange of knowledge from domain specialists to BN modelers and vice versa. This included constructing preliminary models and queries to circulate to prospective participants, to entice participation and stimulate contributions. Our use of independent and intermediary experts with an emphasis on interpreting current literature is unusual, and the total number of expert hours required was exceptionally high.

**Online meetings:** An additional difficulty was that the relevant medical specialists generally have very little time available, and this was further reduced in countries where COVID was most prevalent. Our location in Australia, with its extremely low rates of infection, proved to be fortunate: we were able to use our local contacts to recruit medical specialists who had some time to contribute, and were very generous in donating it. Our common practice is to hold in-person group workshops, but to be COVID-safe, all meetings with external experts were held online. This adaptation proved to have other advantages. Logistically, experts could meet briefly within their busy schedules from their normal places of work in highly dispersed locations without the overhead of travel time, which effectively broadened our pool of available experts and our feasible number of elicitation sessions. Online tools also enabled all our elicitation exchanges to be easily recorded.

**Iterative extension and validation:** Our recruitment and retention was so successful that we were able to follow the kind of iterative, incremental method widely recommended for model building [e.g., Korb and Nicholson, 2011] but rarely realized in expert elicitation. We systematically refined and checked the group output with one-on-one follow-up meetings as necessary. These included both original group members and new, independent recruits, thus incorporating regular expert validation of previous expert input. Participants were invited to “sign off” on all our final elicitation products.

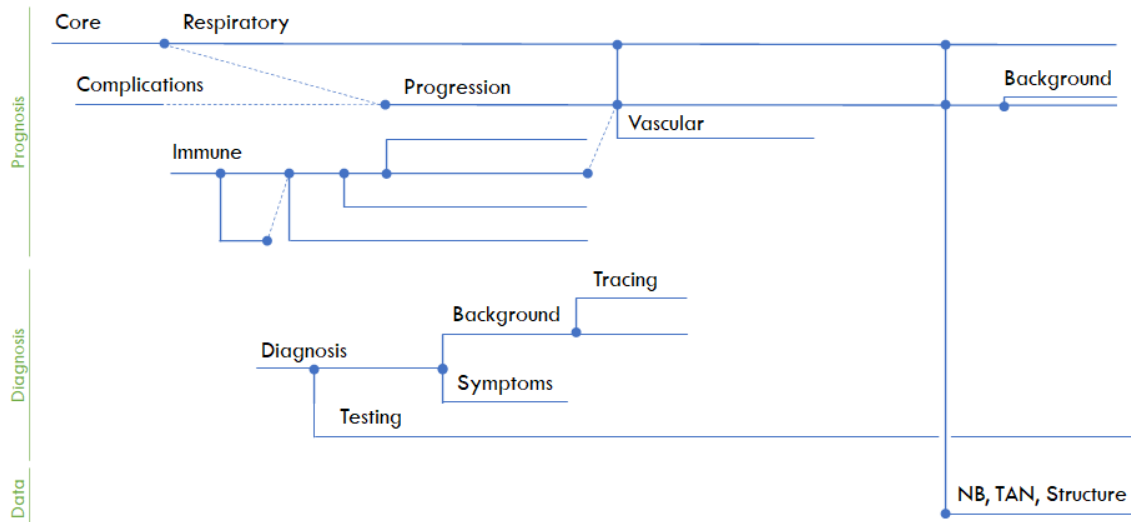


Figure 1: Model cladogram from April to December 2020.

**Documenting theoretical models:** We deliberately aimed for relatively detailed causal models of the domain that captured the experts’ understanding of the relevant processes. We did not avoid discussion and inclusion of theoretically salient latent variables or some other details that might not be useful in the specific, practical BN tools we ultimately aim to develop. Unusually, we will publish separately the full details of, and documentation for, these detailed causal models. This will allow other teams to extend or reparameterize them from new datasets as COVID research continues, and use them to develop or validate their own practical tools.

**Developing practical tools:** Such BN-powered tools often simplify or adapt the theoretical causal structure by including only what is relevant for the purpose, and omitting or merging many latent variables to focus on observable variables that are included in available datasets. They can be readily developed from, or checked against, our theoretical models to ensure that they are as simple as possible while still being consistent with the more detailed theory.

**Parameterizing from patient datasets:** We obtained only indicative qualitative parameters during expert elicitation, in order to help understand and validate the causal structures. Now that substantial patient datasets have accumulated, we are currently deriving precise quantitative parameters and testing our tools for accuracy.

## RESULTS: EVOLUTION OF A MODEL SET

The results of our adaptive research method was not a single model, but rather, an evolving set of models (see Figure 1). There were several reasons for this. Some models were developed with the aim of prognosis, and others primarily for diagnosis. Furthermore, given the complexity of COVID processes, we decided to develop multiple models for differ-

ent physiological subsystems. For example, our Respiratory BN (with 35 variables) primarily described the initial infection process from the airway to the internal organs, whereas our Complications BN modeled the subsequent interactions between these organs. Immune reactions proved complex enough to warrant exploration in their own model, although we subsequently concluded that most of these details were unnecessary for our purposes. Due to the strong feedback loops during COVID complications, the Complications BN mutated into a dynamic BN incorporating some features of the Respiratory BN, dubbed the Progression model.

The project itself has progressed from eliciting general theoretical models from experts to developing our more specific, practically oriented models. The use of alternative databases to parameterize and test models has resulted in further differentiation. In the next phase, we hope to release practical tools that demonstrate the benefits of our comprehensive approach to modeling COVID during lockdown.

## References

- Karl J. Friston, Thomas Parr, Peter Zeidman, Adeel Razi, Guillaume Flandin, Jean Daunizeau, Oliver J. Hulme, Alexander J. Billig, Vladimir Litvak, Rosalyn J. Moran, Cathy J. Price, and Christian Lambert. Dynamic causal modelling of COVID-19, 2020. arXiv:2004.04463.
- Kevin B. Korb and Ann E. Nicholson. *Bayesian Artificial Intelligence*. Chapman & Hall / CRC Press, Boca Raton, 2nd edition, 2011.
- J.G. Richens, C.M. Lee, and S. Johri. Improving the accuracy of medical diagnosis with causal machine learning. *Nature Communications*, 11:3923, 2020.