**Title:** Bridging the gaps in test interpretation of SARS-CoV-2 through Bayesian network modelling

**Authors:** Yue Wu[1], David Foley[2], Jessica Ramsay[3], Owen Woodberry[4], Steven Mascaro[4], Ann E Nicholson[4], Tom Snelling[1,2,3,5,6]

**Affiliations**

1. School of Public Health, University of Sydney, Camperdown, New South Wales, Australia
2. Department of Infectious Diseases, Perth Children's Hospital, Perth, Western Australia, Australia
3. Wesfarmers Centre for Vaccines and Infectious Diseases, Telethon Kids Institute, Perth, Western Australia, Australia
4. Department of Data Science & Artificial Intelligence, Monash University, Clayton, Victoria, Australia
5. School of Public Health, Curtin University, Bentley, Western Australia, Australia
6. Menzies School of Health Research, Charles Darwin University, Darwin, Northern Territory Australia

**Extended abstract**

Bayesian network (BN) models have been used within healthcare to bring clarity to complex problems.[1] In the absence of an established gold standard, an understanding of the testing cycle from individual exposure to test outcome report is required to guide the correct interpretation of SARS-CoV-2 (COVID-19) reverse transcriptase real-time polymerase chain reaction (RT-PCR) results and optimise the testing processes.[2] A wide variation in rates of false negatives has been reported, ranging from 1.8 to 58%;[3] this variability may be attributable to heterogeneity in disease prevalence, patient age, timing of testing, type of specimen, other components of the pre-analytical phase and the RT-PCR assay employed across studies.[4] We use the BN modelling approach to construct a comprehensive framework for understanding the real-world predictive value of individual RT-PCR results.

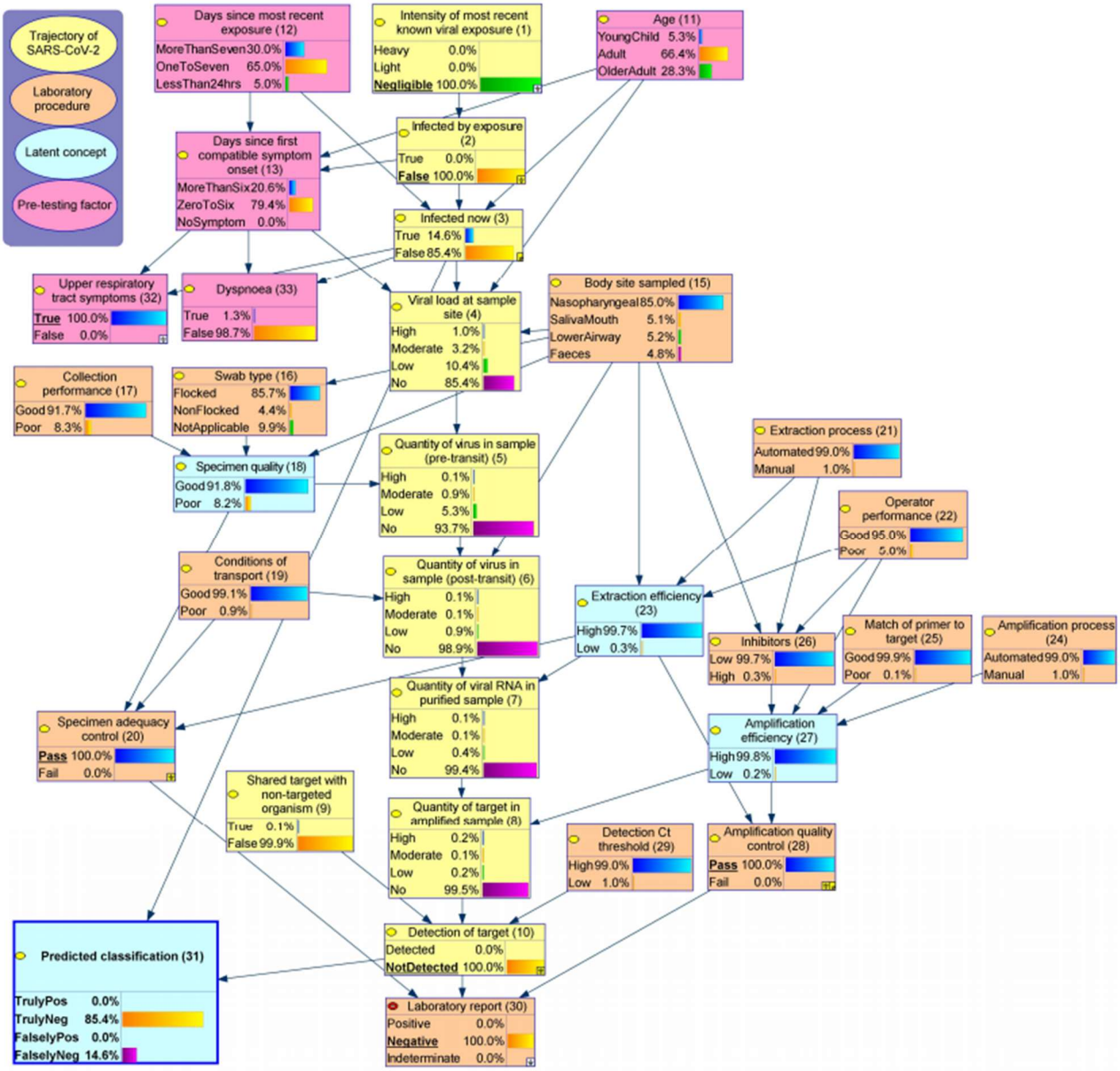
We elicited knowledge from domain experts to describe the test process through a facilitated group workshop. A preliminary model was derived based on the elicited knowledge, then subsequently refined, parameterised and validated with a second workshop and one-on-one discussions. Parameterisation of the model was conducted based on the literature and input from experts.

Causal relationships elicited describe the interactions of pre-testing, specimen collection and laboratory procedures, and RT-PCR platform factors, and their impact on the presence and quantity of virus thus the test result and its interpretation. As shown in Figure 1, the outcome model consists of 32 nodes in total, including 5 pre-testing factors (pink), 13 variables relevant to laboratory procedures (orange), 10 nodes (yellow) sequentially describing the trajectory of SARS-CoV-2 virus from the initial exposure event to final laboratory detection, and 4 summary nodes (blue) created for modelling purpose. 20 variables can be directly mapped to real-world observations therefore used as input variables if available, and the rest 12 variables are latent including the main output variable 'predicted classification' (node ***31***) which describes the probability of the classification being *truly positive*, *truly negative*, *falsely positive*, and *falsely negative*.

By setting the input variables as 'evidence' for a given subject, four scenarios were simulated to demonstrate potential uses of the model. In Figure 1, some input variables were entered to illustrate a scenario where a tested individual who has **negligible known recent viral exposure** (node ***1***) and experiences **upper respiratory symptoms** (node ***32***) but **tested negative** (node ***30***), the probability of a *falsely negative* result is 14.6% (node ***31***) if the person resides in a high viral prevalence setting (5%). The corresponding probability drops to 0.3% if it were a low viral prevalence setting (0.1%).

The core value of this model is a deep understanding of the total testing cycle, bridging the gap between a person's true infection status and their test outcome. This model can be adapted to different settings, testing modalities and pathogens, adding much needed nuance to the interpretations of results. This model requires validation with local, real-world datasets prior to application. We intend for future applications to integrate with other models that detail local epidemiological factors such as those developed by Fenton *et al*.[5] to account for the complex and dynamic interactions between individual-level factors and population-level behaviours that influence the transmission and prevalence of SARS-CoV-2.

***Figure 1. The testing model***




## References


1. McLachlan S, Dube K, Hitman GA, Fenton N, Kyrimi E. Bayesian Networks in Healthcare: Distribution by Medical Condition. *Artificial Intelligence in Medicine.* 2020:101912.
2. Woloshin S, Patel N, Kesselheim AS. False Negative Tests for SARS-CoV-2 Infection—Challenges and Implications. *New England Journal of Medicine.* 2020.
3. Arevalo-Rodriguez I, Buitrago-Garcia D, Simancas-Racines D, et al. False-negative results of initial RT-PCR assays for COVID-19: a systematic review. *medRxiv.* 2020.
4. Kucirka LM, Lauer SA, Laeyendecker O, Boon D, Lessler J. Variation in False-Negative Rate of Reverse Transcriptase Polymerase Chain Reaction-Based SARS-CoV-2 Tests by Time Since Exposure. *Ann Intern Med.* 2020;173(4):262-267.
5. Fenton NE, Neil M, Osman M, McLachlan S. COVID-19 infection and death rates: the need to incorporate causal explanations for the data and avoid bias in testing. *Journal of Risk Research.* 2020;23(7-8):862-865.