DE GRUYTER
Rev Environ Health 2016; 31(1): 173–177

Colleen L. Lau* and Carl S. Smith

Bayesian networks in infectious disease eco-epidemiology

DOI 10.1515/reveh-2015-0052
Received October 15, 2015; accepted October 16, 2015; previously published online January 20, 2016

Abstract: Globally, infectious diseases are responsible for a significant burden on human health. Drivers of disease transmission depend on interactions between humans, the environment, vectors, carriers, and pathogens; transmission dynamics are therefore potentially highly complex. Research in infectious disease eco-epidemiology has been rapidly gaining momentum because of the rising global importance of disease emergence and outbreaks, and growing understanding of the intimate links between human health and the environment. The scientific community is increasingly recognising the need for multidisciplinary translational research, integrated approaches, and innovative methods and tools to optimise risk prediction and control measures. Environmental health experts have also identified the need for more advanced analytical and biostatistical approaches to better determine causality, and deal with unknowns and uncertainties inherent in complex systems. In this paper, we discuss the use of Bayesian networks in infectious disease eco-epidemiology, and the potential for developing dynamic tools for public health decision-making and improving intervention strategies.

Keywords: Bayesian networks; eco-epidemiology; infectious disease epidemiology; leptospirosis; zoonoses.

Eco-epidemiology of infectious diseases – global significance

Globally, infectious diseases are responsible for a significant burden on human health. Drivers of disease transmission depend on interactions between humans, the natural environment (e.g. climate, flooding, natural disasters), the anthropogenic environment (e.g. sanitation, hygiene, urbanisation, agriculture, resource development), vectors (e.g. animals, insects), carriers (e.g. water, soil, air), and pathogens. An estimated 60% of emerging pathogens are of zoonotic origin (1), highlighting the importance of interactions between humans, animals, and the environment in driving disease emergence (2). Zoonotic disease transmission is highly complex, and could vary significantly between environmental settings. For example, leptospirosis is the most common bacterial zoonosis worldwide, with >500,000 severe infections annually, and up to 30% case-fatality (3). Leptospirosis is an emerging infectious disease in many contexts, with increasing frequency and severity of outbreaks; changing geographic distribution and serovar patterns; and evolving climatic, socio-demographic, and environmental drivers of transmission (4–6). Infection is acquired through contact with animals, or contaminated soil and water. Risk factors for transmission are highly complex, and vary between countries (4, 7, 8). In Australia, cases are mostly related to occupational exposure in farmers and recreational exposure to freshwater (7, 9). In India, large outbreaks occur each year in rodent-infested urban slums during monsoonal flooding that spreads bacteria and simultaneously displaces humans and rodents (4). In American Samoa, leptospirosis infections have been linked to human behavioural factors, backyard piggeries, urbanisation, land use, population density, and flooding risk (10–12). The World Health Organisation has identified leptospirosis as a neglected tropical disease of zoonotic origin, with large knowledge gaps in epidemiology, disease burden, transmission dynamics, and evidence-based interventions. The ability to plan, implement and sustain effective control measures and mitigation strategies have been restricted by a lack of epidemiological data and effective tools to accurately identify, predict, and forecast outbreaks (3).

***Corresponding author: Colleen L. Lau,** Children's Health and Environment Program, L7 Child Health Research Centre, The University of Queensland, 62 Graham St, South Brisbane, Queensland 4101, Australia; and Queensland Children's Medical Research Institute, Centre for Children's Health Research, Brisbane, Australia, E-mail: colleen.lau@uq.edu.au
Carl S. Smith: School of Agriculture and Food Sciences, The University of Queensland, Australia

Eco-epidemiology research: need for innovative approaches

Research in infectious disease eco-epidemiology has been rapidly gaining momentum because of the rising

global importance of disease emergence and outbreaks, and growing understanding of the intimate links between human health and the environment. The scientific community is increasingly recognising the need for multidisciplinary translational research and integrated approaches required to understand risk factors and exposure pathways, inform public health mitigation and interventions, and develop innovative methods and tools to optimise predictions and control (13, 14). Environmental health experts have also identified the need for more advanced analytical and biostatistical approaches to better determine causality, and deal with unknowns and uncertainties inherent in complex systems (15, 16).

Limitations of regression models

Regression analysis is the most common approach used to assess risk factors in epidemiology. However, statistical associations between risk factors and outcomes do not provide any insights into cause and effect, or explain exposure pathways. Regression models also usually assume that risk factors are independent of each other, and could result in oversimplified models of complex systems. Unless care and rigour are applied to assess interactions between variables and interpret these effects, important risk factors might be 'dropped out' of models because of confounding correlations, and result in lost opportunities for mitigation and control. Regression models are limited in their ability to incorporate unknowns or uncertainties, and have poor capacity to disentangle the intricate associations between risk factors, drivers, triggers, and outcomes (15, 17, 18). Also, regression models have limited ability to predict events for which there are no historical data, or catastrophic events such as unprecedented outbreaks which are 'off the scale' of models (19). Furthermore, regression models cannot be rapidly updated without high levels of technical expertise – a significant disadvantage if they are to be used as tools for risk factor analysis and decision-making during an outbreak or disaster.

Bayesian networks

Bayesian networks (also called Bayesian belief networks) provide a much more powerful platform to understand interrelations between components of complex systems, determine cause and effect, assess risks and opportunities, and assist with decision-making (17–19). BNs have therefore been widely used in many disciplines including medicine, ecology, agriculture, environmental science, engineering, economics, chess, and artificial intelligence (20–22). In medicine, BNs and conditional probabilities are used for diagnosis, treatment, and prognosis of conditions including cancers and cardiovascular disease, decision support systems for patient care, and understanding cellular-level molecular mechanisms (23–26). Application of BNs in infectious disease eco-epidemiology and public health has so far been extremely limited.

BNs are composed of two main components: i) directed acyclic graphs (DAGs) which consist of nodes that represent variables and outcomes, and arrows that define the causal links between nodes, and ii) conditional probability tables for each node that define quantitative relationships between nodes (17). DAGs provide a visual, explicit, and easily interpreted representation of causal pathways, hypotheses, and assumptions in complex systems. Variables (nodes) are categorised into states (e.g. yes/no), and relationships between nodes are based on the Bayes theorem of conditional probability. The probability of states in a node depends on the conditional probabilities of all nodes that feed into it. When a scenario is inserted into a BN, they are able to update probabilities from cause to effect, and also from effect to cause, using both forward and backward propagation. BNs are therefore able to characterise both magnitude and direction of associations between variables (19).

A very simple example of a network for a leptospirosis study might include three nodes: node1=sex, states=male or female; node2=ethnic group, states=indigenous or other; node3=leptospirosis infection, states=not infected or infected (Figure 1A–C). If we believe that sex and ethnicity are risk factors for leptospirosis, direct links can be created from node1 to node3, and from node2 to node3. Figure 1A shows the network containing probabilities learnt from a theoretical database of study participants, with probabilities for states in each node reflecting observations in the data. By selecting the states of male in node1 and indigenous in node2, the BN updates to show the probability of leptospirosis infection in indigenous males (Figure 1B). Similarly, different probabilities can be determined for each combination of sex and ethnic group. Alternatively, if infected is selected in node3 the BN updates the probabilities for node1 and node2 to show the distribution of probabilities across sex and ethnic group for infected people (Figure 1C).

In complex models (e.g. Figure 2), BNs are able to calculate probabilities associated with any event or scenario, provide a visual portal for testing subjective probabilities

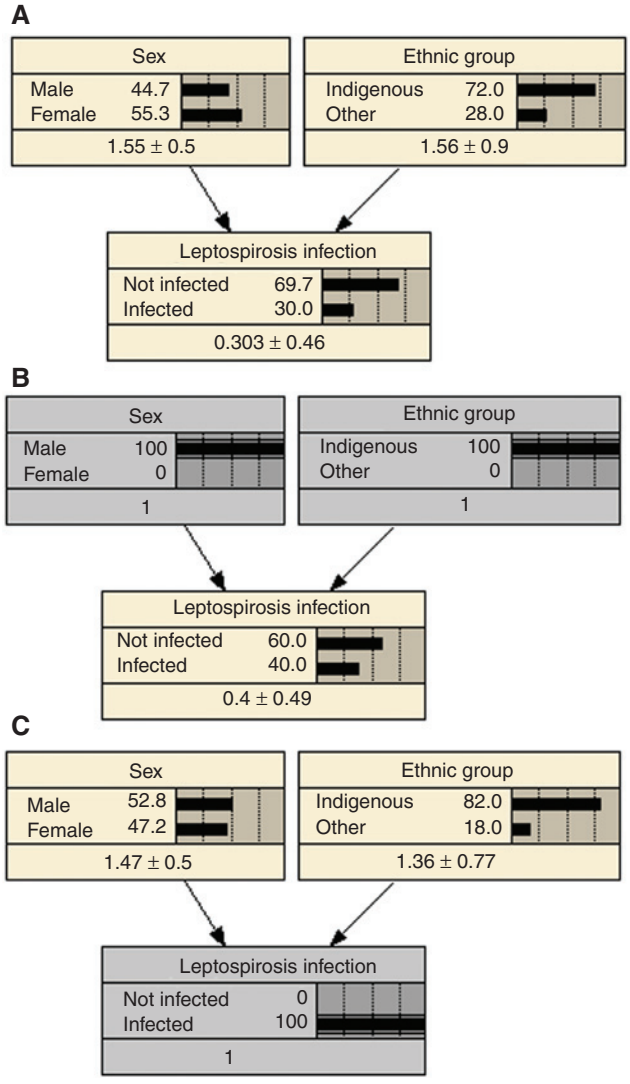

Figure 1: Simple example of a Bayesian network for a leptospirosis study with three nodes: node1=sex, states=male or female; node2=ethnic group, states=indigenous or other; node3=leptospirosis infection, states=not infected or infected (Figure 1A–C).

(or expert opinion) when there are no data, and for defining the range of outcomes with best/worst case scenarios. Risk measures are more meaningful and more easily put into context when visualised as causal pathways that explain potential sources of pathogens, routes of exposure, and drivers of transmission. When triggers and control measures are built into the network as nodes, BNs can also be used to quantify the impact of actions or inactions. The user-friendly and visually interactive interface provided by BN software such as Netica (Norsys Software Corp, Vancouver, Canada) allows scenarios to be entered into the network with the click of a mouse and effective demonstration of results to stakeholders even if they have little understanding of BNs or biostatistics.

Types of BNs include naïve networks (in which an outcome is the direct parent of several predictor variables and there or no causal links between the predictor variables), expert-structured networks such as Figure 2 (causal links between variables and outcomes are determined by knowledge and beliefs), and machine-structured networks (only variables are defined and links variables are automatically generated based on maximum likelihood). Performance and predictive accuracy of models can be compared using robust statistical measures including sensitivity, specificity, true skills statistic (for Boolean outcomes), KAPPA statistic (for outcomes with more than two states), classification error rates, and ROC curves (for Boolean outcomes).

Causal models such as BNs allow us to dig deeper into data to explore exposure pathways, gain insights into how best to control risk, and determine scenarios that will maximise the probability of desired outcomes (17, 19). Complex models could incorporate interrelationships between multiple causes, consequences, and exposure pathways, and therefore more closely reflect infectious disease transmission in the real world. BNs calculate the probability of events based on current knowledge, are easily updated as more information becomes available, and outputs are computed quickly and efficiently even for large models. Expert opinion can be used to set prior and condition probabilities for variables for which we have no data, and updated as beliefs or hypotheses evolve. BNs can incorporate not only risk events and consequences, but also trigger factors and control measures that might include primary, secondary, and tertiary interventions. Hypotheses regarding risk factors and causal pathways can be tested by building alternative model structures, e.g. reversing the direction of causal links between variables, removing/adding links, and comparing quantitative measures of model performance (19).

Incorporating geospatial components into Bayesian networks

Spatial epidemiology is increasingly being used in infectious disease research to define disease distribution, elucidate drivers of transmission, and produce maps to inform decisions (27, 28). Predictive risk maps have been produced for a wide range of infectious diseases including dengue, malaria, helminths, and leptospirosis (11, 28–30).

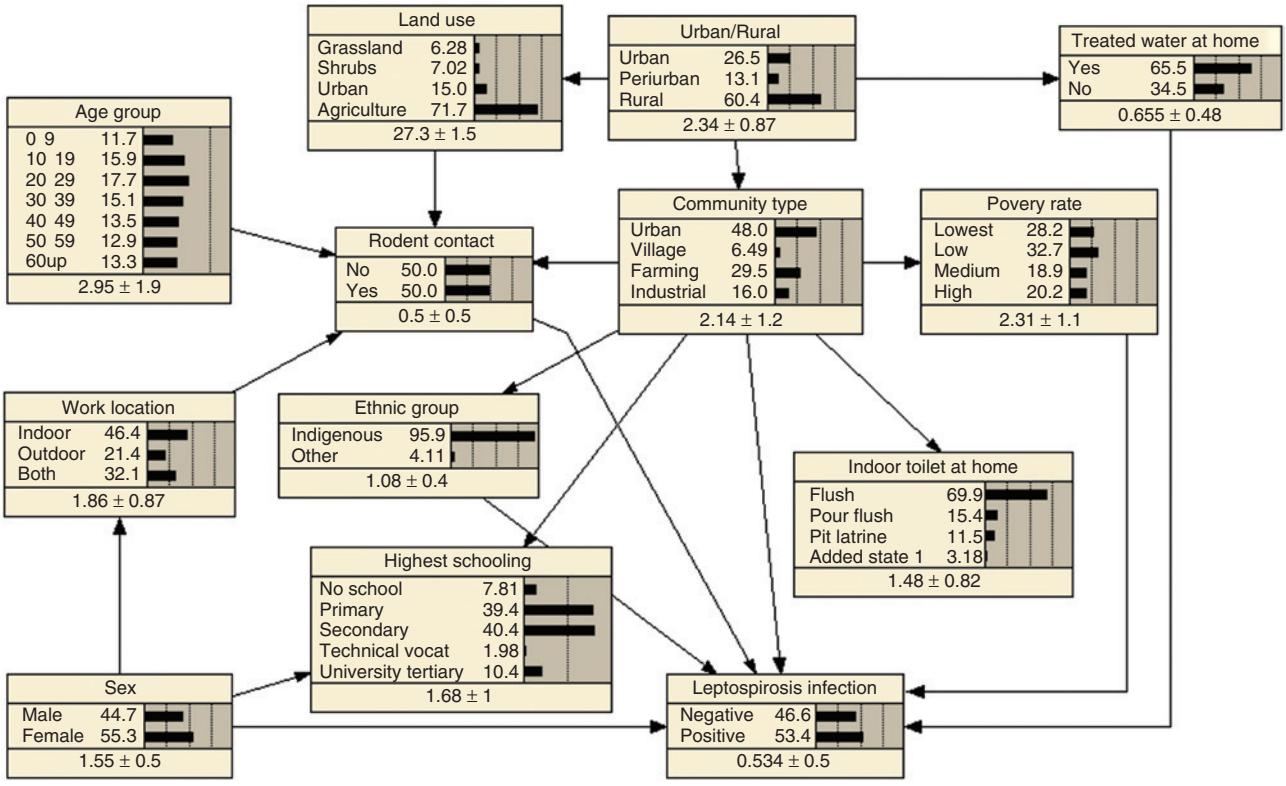

Figure 2: Example of a more complex Bayesian network for assessing the risk of leptospirosis using a theoretical database.
The BN includes multiple risk factors for leptospirosis infection, interrelationships between risk factors, and potential causal pathways. The probability distributions shown were learnt from a theoretical database.

A recent development in Bayesian network modelling is software such as GeoNetica (Norsys Software Corp, Vancouver, Canada) that seamlessly integrates geographic information systems (GIS) and BNs, and provides powerful platforms for incorporating geospatial inputs and producing geospatial outputs. Geospatial inputs could include any variables that affect disease transmission, e.g. climate, topography, land use, population density, socio-economic and census data, and access to health services. Geospatial outputs such as dynamic disease risk maps can be generated very quickly, and changes in disease risk distribution and hotspots demonstrated for different scenarios (e.g. flooding). Simultaneous outputs of BNs and maps provide a more comprehensive picture of risk factors, causal pathways, triggers, drivers, hotspots, and control measures. This capability has important benefits for real-world public health intelligence, such as 'on-the-fly' hotspot mapping of disease emergence and outbreaks. If scenarios are changed, maps will be updated rapidly to produce visually powerfully outputs. As new geospatial data become available (e.g. new census, higher resolution land use data), they can also be easily incorporated into existing BNs.

Conclusion

Bayesian networks could improve our understanding of the eco-epidemiology of infectious diseases, especially those with strong environmental drivers of transmission. The interactive nature of BNs and the ability to incorporate spatial components and rapidly update predictions also provides potential for developing dynamic tools for public health decision-making and improving intervention strategies.

References

1. Jones KE, Patel NG, Levy MA, Storeygard A, Balk D, et al. Global trends in emerging infectious diseases. Nature 2008;451:990–3.
2. Dixon MA, Dar OA, Heymann DL. Emerging infectious diseases: opportunities at the human-animal-environment interface. Vet Rec 2014;174:546–51.
3. World Health Organization. Report of the Second Meeting of the Leptospirosis Burden Epidmiology Reference Group. 2011. Available at: http://whqlibdoc.who.int/publications/2011/9789241501521_eng.pdf. Accessed on 23 September 2015.

4. Lau CL, Smythe LD, Craig SB, Weinstein P. Climate change, flooding, urbanisation and leptospirosis: fuelling the fire? Trans R Soc Trop Med Hyg 2010;104:631–8.
5. Derne BT, Fearnley EJ, Lau CL, Paynter S, Weinstein P. Biodiversity and leptospirosis risk: a case of pathogen regulation? Med Hypoth 2011;77:339–44.
6. Hartskeerl RA, Collares-Pereira M, Ellis WA. Emergence, control and re-emerging leptospirosis: dynamics of infection in the changing world. Clin Microbiol Infect 2011;17:494–501.
7. Lau C, Smythe L, Weinstein P. Leptospirosis: an emerging disease in travellers. Travel Med Infect Dis 2010;8:33–9.
8. Bharti AR, Nally JE, Ricaldi JN, Matthias MA, Diaz MM, et al. Leptospirosis: a zoonotic disease of global importance. Lancet Infect Dis 2003;3:757–71.
9. Lau CL, Skelly C, Dohnt M, Smythe LD. The Emergence of *Leptospira Borgpetersenii* Serovar Arborea in Queensland, Australia, 2001 to 2013. BMC Infect Dis 2015;15:230.
10. Lau CL, Skelly C, Smythe LD, Craig SB, Weinstein P. Emergence of new leptospiral serovars in American Samoa – ascertainment or ecological change? BMC Infect Dis 2012;12:19.
11. Lau CL, Clements AC, Skelly C, Dobson AJ, Smythe LD, et al. Leptospirosis in American Samoa – estimating and mapping risk using environmental data. PLoS Negl Trop Dis 2012;6:e1669.
12. Lau CL, Dobson AJ, Smythe LD, Fearnley EJ, Skelly C, et al. Leptospirosis in American Samoa 2010: epidemiology, environmental drivers, and the management of emergence. Am J Trop Med Hyg 2012;86:309–19.
13. Rabinowitz PM, Kock R, Kachani M, Kunkel R, Thomas J, et al. Toward proof of concept of a one health approach to disease prediction and control. Emerg Infect Dis 2013;19:e130265.
14. Coker R, Rushton J, Mounier-Jack S, Karimuribo E, Lutumba P, et al. Towards a Conceptual Framework to Support One-Health Research for Policy on Emerging Zoonoses. Lancet Infect Dis 2011;11:326–31.
15. Burns CJ, Wright JM, Pierson JB, Bateson TF, Burstyn I, et al. Evaluating uncertainty to strengthen epidemiologic data for use in human health risk assessments. Environ Health Perspect 2014;122:1160–5.
16. Eisenberg JN, Desai MA, Levy K, Bates SJ, Liang S, et al. Environmental determinants of infectious disease: a framework for tracking causal links and guiding public health research. Environ Health Perspect 2007;115:1216–23.
17. Joffe M, Gambhir M, Chadeau-Hyam M, Vineis P. Causal diagrams in systems epidemiology. Emerg Themes Epidemiol 2012;9:1.
18. Landuyt D, Broekx S, D'Hondt R, Engelen G, Aertsens J, et al. A Review of Bayesian Belief Networks in Ecosystem Service Modelling. Environ Model Softw 2013;46:1–11.
19. Fenton N, Neil M. Risk assessment and decision analysis with bayesian networks. Boca Raton: CRC Press, Taylor & Francis Group, 2013.
20. Bashari H, Smith C, Bosch OJ. Developing decision support tools for rangeland management by combining state and transition models and Bayesian belief networks. Agri Syst 2008;99:23–4.
21. Smith CS, Howes AL, Price B, McAlpine CA. Using a Bayesian belief network to predict suitable habitat of an endangered mammal – the Julia Creek dunnart (Sminthopsis douglasi). Biologic Conser 2007;139:333–47.
22. Liedloff AC, Smith CS. Predicting a 'tree change' in Australia's tropical savannas: combining different types of models to understand complex ecosystem behaviour. Ecol Model 2010;221:2565–75.
23. Gevaert O, De Smet F, Timmerman D, Moreau Y, De Moor B. Predicting the prognosis of breast cancer by integrating clinical and microarray data with Bayesian networks. Bioinformatics 2006;22:e184–90.
24. Sesen MB, Nicholson AE, Banares-Alcantara R, Kadir T, Brady M. Bayesian networks for clinical decision support in lung cancer care. PLoS One 2013;8:e82349.
25. Lucas PJ, van der Gaag LC, Abu-Hanna A. Bayesian networks in biomedicine and health-care. Artif Intell Med 2004;30:201–14.
26. Wang KJ, Makond B, Wang KM. Modeling and predicting the occurrence of brain metastasis from lung cancer by Bayesian network: a case study of Taiwan. Comput Biol Med 2014;47:147–60.
27. Hay SI, Battle KE, Pigott DM, Smith DL, Moyes CL, et al. Global mapping of infectious disease. Philos Trans R Soc Lond B Biol Sci 2013;368:20120250.
28. Clements AC, Reid HL, Kelly GC, Hay SI. Further shrinking the malaria map: how can geospatial science help to achieve malaria elimination? Lancet Infect Dis 2013;13:709–18.
29. Bhatt S, Gething PW, Brady OJ, Messina JP, Farlow AW, et al. The global distribution and burden of dengue. Nature 2013;496:504–7.
30. Magalhaes RJ, Clements AC, Patil AP, Gething PW, Brooker S. The applications of model-based geostatistics in helminth epidemiology and control. Adv Parasitol 2011;74:267–96.