Environmental Modelling & Software 97 (2017) 271–286

Contents lists available at ScienceDirect

Environmental Modelling & Software

journal homepage: www.elsevier.com/locate/envsoft

Unravelling infectious disease eco-epidemiology using Bayesian networks and scenario analysis: A case study of leptospirosis in Fiji

Colleen L. Lau [a, b, *], Helen J. Mayfield [a], John H. Lowry [c, d], Conall H. Watson [e], Mike Kama [f], Eric J. Nilles [g], Carl S. Smith [h]

[a] Research School of Population Health, Australian National University, Canberra, Australian Capital Territory, Australia
[b] Child Health Research Centre, The University of Queensland, Brisbane, Queensland, Australia
[c] School of People, Environment and Planning, Massey University, Palmerston North, New Zealand
[d] School of Geography, Earth Science and Environment, University of the South Pacific, Suva, Fiji
[e] Centre for the Mathematical Modelling of Infectious Diseases, London School of Hygiene & Tropical Medicine, London, United Kingdom
[f] Fiji Centre for Communicable Disease Control, Ministry of Health and Medical Services, Suva, Fiji
[g] Division of Pacific Technical Support, World Health Organization, Suva, Fiji
[h] School of Business, University of Queensland, Brisbane, Queensland, Australia

ARTICLE INFO

Article history:
Received 21 February 2017
Received in revised form
8 August 2017
Accepted 9 August 2017
Available online 8 September 2017

Keywords:
Bayesian networks
Infectious diseases epidemiology
Leptospirosis
Zoonoses
Environmental health
Public health

ABSTRACT

Regression models are the standard approaches used in infectious disease epidemiology, but have limited ability to represent causality or complexity. We explore Bayesian networks (BNs) as an alternative approach for modelling infectious disease transmission, using leptospirosis as an example. Data were obtained from a leptospirosis study in Fiji in 2013. We compared the performance of naïve versus expert-structured BNs for modelling the relative importance of animal species in disease transmission in different ethnic groups and residential settings. For BNs of animal exposures at the individual/household level, R^2 for predicted versus observed infection rates were 0.59 for naïve and 0.75–0.93 for structured models of ethnic groups; and 0.54 for naïve and 0.93–1.00 for structured models of residential settings. BNs provide a promising approach for modelling infectious disease transmission under complex scenarios. The relative importance of animal species varied between subgroups, with important implications for more targeted public health control strategies.

Software availability

Name: Netica version 5.12
Developer: Norsys Software Corporation
Address: 3512 West 23rd Ave, Vancouver, BC, Canada
Tel: +1 604 221 2223
Email: info@norsys.com
Availability: www.norsys.com

Data availability

The data were collected from small communities in Fiji, and participants could potentially be re-identifiable if the study data were fully available, e.g. by diagnosis of leptospirosis, demographics, occupation, and household GPS locations. Public deposition of the data would compromise participant privacy, and therefore breach compliance with the protocol approved by the research ethics committees. For researchers who meet the criteria for access to confidential information, data can be requested via the Human Research Ethics Committee at the Australian National University. Email: human.ethics.officer@anu.edu.au. Phone: +61 (2) 6125 3427.

1. Introduction

The growing discipline of infectious disease eco-epidemiology seeks to understand the environmental, ecological, and socio-demographic drivers of emergence, transmission, and outbreaks (Eisenberg et al., 2007; Rabinowitz et al., 2013; Mills et al., 2010). The drivers depend on complex interactions between humans, the natural environment (e.g. climate and vegetation), the

---

* Corresponding author. Department of Global Health, Research School of Population Health, Australian National University, 62 Mills Road, Canberra, ACT 0200, Australia.
E-mail address: colleen.lau@anu.edu.au (C.L. Lau).

http://dx.doi.org/10.1016/j.envsoft.2017.08.004

anthropogenic environment (e.g. urbanisation and land use), vectors (e.g. insects and animals), and carriers (e.g. water, soil, and air) (Jones et al., 2008). Regression models are the most common approaches to risk factor analysis in infectious disease epidemiology; while they are widely accepted and understood, there are important drawbacks when studying complex systems, and the need for more novel epidemiological approaches are being increasingly recognised (McMichael, 1999; Susser, 1998; Burns et al., 2014; Lau et al., 2016a; Keyes et al., 2017; Poole, 2017). Standard regression models rely on an explicit assumption of independence amongst the predictor variables as well as independence between units, which is often not true in the real world of disease transmission, and could potentially result in oversimplification of models. Standard regression models do not allow strongly correlated predictor variables to be retained, even if each variable might play crucial and distinct roles in transmission. Standard regression models therefore have limitations in their capacity to disentangle the intricate associations between risk factors, drivers, triggers, and outcomes (Burns et al., 2014).

Causal models such as Bayesian networks (BNs) have the ability to represent causality as well as incorporate relationships between predictor/indicator variables, and may provide an alternative approach to more accurately model complex systems (Joffe et al., 2012; Landuyt et al., 2013). Other methods used to model complex systems and incorporate collinearity include the use of interactions in regression analysis, regression trees, structured equation models, path analysis and multilevel hierarchical models. Compared to these methods, Bayesian network models have added advantages of being both visually more intuitive and having interactive interfaces that can be used to assess complex scenarios and produce real-time outputs. In particular, the ability to define scenarios that include strongly correlated predictor variables is difficult to achieve with regression models. However, BNs also have certain limitations when modelling complex systems. BNs generally use discretised variables and produced outputs that are discrete outcomes or events, and discretisation of continuous variables is sometimes associated with loss of data resolution. Also, BNs are not dynamic and cannot incorporate feedback loops, a potentially important consideration for complex models.

Leptospirosis is an important zoonotic disease worldwide that causes an estimated one million severe cases per year, with particularly high risk in tropical and subtropical regions (Costa et al., 2015; Bharti et al., 2003). Humans are infected through direct contact with infected mammals (including rodents, livestock, pets, and wildlife), or contact with water or soil contaminated by urine of infected animals. Drivers of transmission are complex and include individual behaviour, socio-demographics, culture, lifestyle, contact with animals, and the natural environment (Lau et al., 2010; Mwachui et al., 2015; Jones et al., 2013). Environmental drivers for leptospirosis transmission, emergence, and outbreaks are increasingly being recognised, raising concerns that transmission and flood-related outbreaks could intensify with global change in both natural and anthropogenic environments (Lau et al., 2010; Hartskeerl et al., 2011; Coker et al., 2011). In developing countries, rapid population growth often results in urbanisation, slums, poor sanitation, poverty, subsistence livestock and agricultural intensification – all of which are important drivers of zoonotic disease transmission (Jones et al., 2013; Liverani et al., 2013). The Pacific Islands are particularly vulnerable to the health impacts of climate change because of all of the socio-demographic, geographic, and environmental factors mentioned above (McIver et al., 2015; Averett, 2016), and leptospirosis causes significant health impact in the region (Kline et al., 2013; Weinberger et al., 2014; Massenet et al., 2015; Colt et al., 2014; Lau et al, 2012, 2016b).

Over the past decades, Fiji has experienced increasing incidence and outbreaks of leptospirosis (Lau et al., 2016b; Ram et al., 1982; Ghosh et al., 2010; McIver et al., 2012). Two post-flooding outbreaks occurred in 2012, resulting in over 500 cases and 40 deaths. An eco-epidemiological study conducted in 2013 found a community leptospirosis seroprevalence (the percentage of a population with detectable leptospirosis antibodies in their blood) of 19·4% using the microscopic agglutination test (MAT), with significant variation between ethnic groups and residential settings. The findings of the study have been published, focusing on risk factor analysis using standard regression approaches (Lau et al., 2016b). The study provided important insights into leptospirosis eco-epidemiology in Fiji, but there remain multiple unanswered questions with important public health implications. Important questions regarding the reasons for the disparate risk between ethnic groups and residential settings have not been clearly answered, but it is possible that niche-specific interventions may be required for more effective public health control measures. For example, intervention strategies may need a different focus for each ethnic group and/or vary between urban, peri-urban, and rural areas. The study also raised questions about the relative importance of animal species in human infections, a fundamental question when prioritising public health interventions for leptospirosis. On univariate regression analysis, infection was associated with contact with multiple animal species, including rodents, mongoose, dogs, and multiple species of livestock. However, there were significant correlations between presence of different animals species (e.g. people who own pigs are also more likely to own cows), and on multivariable regression analyses, the only animal-related predictor variables retained in the final model were the presence of pigs in the community and high cattle density. Based on these results, can we assume that animal species other than pigs and cattle did not play an important role in human infections? Or could other species be also important, but excluded from multivariable regression models because they were strongly correlated with exposure to pigs or cattle? Also, might the relative importance of different animal species differ between ethnic groups and residential settings, and therefore require more tailored interventions? These questions highlight some of the limitations of using standard regression analysis to model infectious diseases with complex transmission dynamics and environmental drivers.

In this paper, we explore the use of BNs as an alternative methodological approach for modelling the eco-epidemiology of infectious diseases, using leptospirosis in Fiji as a case study. Firstly, the study aims to improve model performance of BNs by building models that better represent and explain causality. Secondly, the study aims to use BNs to determine the relative importance of animal species in disease transmission in different ethnic groups and residential settings.

2. Materials and methods

2.1. Study location and setting

Fiji has a population of 837,217 (Fiji Bureau of Statistics, 2007) living in urban, peri-urban, and rural settings in tropical islands. Two main ethnic groups, iTaukei (indigenous Fijian) and Indo-Fijians (Fijians of Indian descent), account for 57% and 35% of the population respectively (Fiji Bureau of Statistics, 2007). Subsistence livestock are common in backyards and communal areas, particularly in rural areas. Rodents, mongoose, dogs, and cats are abundant in both urban and rural areas.

2.2. Data sources

This study used a database from a recently published study of