

1
2
3
4 **Title: Unravelling Infectious Disease Eco-epidemiology using Bayesian Networks and**
5 **Scenario Analysis: A Case Study of Leptospirosis in Fiji**
6
7
8
9

10 **Authors and affiliations:**

11
12 Colleen Lau^{a,b}, Helen Mayfield^a, John Lowry^{c,d}, Conall Watson^e, Mike Kama^f, Eric Nilles^g, Carl
13
14 Smith^h

15
16 ^a Department of Global Health, Research School of Population Health, Australian National
17
18 University, Canberra, ACT, Australia

19 ^b Child Health Research Centre, The University of Queensland, Brisbane, QLD, Australia

20
21 ^c School of People, Environment and Planning, Massey University, Palmerston North, New
22
23 Zealand

24 ^d School of Geography, Earth Science and Environment, University of the South Pacific, Suva, Fiji

25
26 ^e Centre for the Mathematical Modelling of Infectious Diseases, London School of Hygiene &
27
28 Tropical Medicine, London, United Kingdom

29 ^f Fiji Centre for Communicable Disease Control, Ministry of Health and Medical Services, Suva,
30
31 Fiji

32 ^g Division of Pacific Technical Support, World Health Organization, Suva, Fiji

33
34 ^h School of Agriculture and Food Science, University of Queensland, Brisbane, Queensland,
35
36 Australia

37
38 **Corresponding author:**

39
40 Dr Colleen Lau (MBBS, MPH&TM, PhD, FRACGP)

41
42 Department of Global Health, Research School of Population Health,
43
44 Australian National University

45
46 62 Mills Road, Canberra, ACT 0200, Australia

47
48 Email: colleen.lau@anu.edu.au. Phone: +61 402134878. Fax: +61 2 6125 5608.
49
50
51
52
53
54
55
56
57
58
59

60
61
62 **ABSTRACT**
63

64 Regression models are the standard approaches used in infectious disease epidemiology, but have
65 limited ability to represent causality or complexity. We explore Bayesian networks (BNs) as an
66 alternative approach for modelling infectious disease transmission, using leptospirosis as an
67 example. Data were obtained from a leptospirosis study in Fiji in 2013. We compared the
68 performance of naïve versus expert-structured BNs for modelling the relative importance of animal
69 species in disease transmission in different ethnic groups and residential settings. For BNs of animal
70 exposures at the individual/household level, R^2 for predicted versus observed infection rates were
71 0.59 for naïve and 0.75-0.93 for structured models of ethnic groups; and 0.54 for naïve and 0.93-
72 1.00 for structured models of residential settings. BNs provide a promising approach for modelling
73 infectious disease transmission under complex scenarios. The relative importance of animal species
74 varied between subgroups, with important implications for more targeted public health control
75 strategies.
76
77
78
79
80
81
82
83
84

85 **KEYWORDS**
86

87 Bayesian Networks, Infectious Diseases Epidemiology, Leptospirosis, Zoonoses, Environmental
88 Health, Public Health
89
90

91 **SOFTWARE AVAILABILITY**
92

93 Name: Netica version 5.12
94

95 Developer: Norsys Software Corporation
96

97 Address: 3512 West 23rd Ave, Vancouver, BC, Canada
98

99 Tel: +1 604 221 2223. Email: info@norsys.com
100

101 Availability: www.norsys.com
102

103 **DATA AVAILABILITY:**
104

105 The data were collected from small communities in Fiji, and participants could potentially be re-
106 identifiable if the study data were fully available, e.g. by diagnosis of leptospirosis, demographics,
107 occupation, and household GPS locations. Public deposition of the data would compromise
108 participant privacy, and therefore breach compliance with the protocol approved by the research
109 ethics committees. For researchers who meet the criteria for access to confidential information, data
110 can be requested via the Human Research Ethics Committee at the Australian National University.
111 Email: human.ethics.officer@anu.edu.au. Phone: +61 (2) 6125 3427.
112
113
114
115
116
117
118

119
120
121 **INTRODUCTION**
122
123

124 The growing discipline of infectious disease eco-epidemiology seeks to understand the
125 environmental, ecological, and socio-demographic drivers of emergence, transmission, and
126 outbreaks.¹⁻³ The drivers depend on complex interactions between humans, the natural environment
127 (e.g. climate and vegetation), the anthropogenic environment (e.g. urbanisation and land use),
128 vectors (e.g. insects and animals), and carriers (e.g. water, soil, and air).⁴ Regression models are the
129 most common approaches to risk factor analysis in infectious disease epidemiology; while they are
130 widely accepted and understood, there are important drawbacks when studying complex systems,
131 and the need for more novel epidemiological approaches are being increasingly recognised.⁵⁻¹⁰
132 Standard regression models rely on an explicit assumption of independence amongst the predictor
133 variables as well as independence between units, which is often not true in the real world of disease
134 transmission, and could potentially result in oversimplification of models. Standard regression
135 models do not allow strongly correlated predictor variables to be retained, even if each variable
136 might play crucial and distinct roles in transmission. Standard regression models therefore have
137 limitations in their capacity to disentangle the intricate associations between risk factors, drivers,
138 triggers, and outcomes.⁷
139

140 Causal models such as Bayesian networks (BNs) have the ability to represent causality as well as
141 incorporate relationships between predictor/indicator variables, and may provide an alternative
142 approach to more accurately model complex systems.^{11,12} Other methods used to model complex
143 systems and incorporate collinearity include the use of interactions in regression analysis,
144 regression trees, structured equation models, path analysis and multilevel hierarchical models.
145 Compared to these methods, Bayesian network models have added advantages of being both
146 visually more intuitive and having interactive interfaces that can be used to assess complex
147 scenarios and produce real-time outputs. In particular, the ability to define scenarios that include
148 strongly correlated predictor variables is difficult to achieve with regression models. However, BNs
149 also have certain limitations when modelling complex systems. BNs generally use discretised
150 variables and produced outputs that are discrete outcomes or events, and discretisation of
151 continuous variables is sometimes associated with loss of data resolution. Also, BNs are not
152 dynamic and cannot incorporate feedback loops, a potentially important consideration for complex
153 models.
154

155 Leptospirosis is an important zoonotic disease worldwide that causes an estimated one million
156 severe cases per year, with particularly high risk in tropical and subtropical regions.^{13,14} Humans
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177

178
179
180 are infected through direct contact with infected mammals (including rodents, livestock, pets, and
181 wildlife), or contact with water or soil contaminated by urine of infected animals. Drivers of
182 transmission are complex and include individual behaviour, socio-demographics, culture, lifestyle,
183 contact with animals, and the natural environment.¹⁵⁻¹⁷ Environmental drivers for leptospirosis
184 transmission, emergence, and outbreaks are increasingly being recognised, raising concerns that
185 transmission and flood-related outbreaks could intensify with global change in both natural and
186 anthropogenic environments.^{15,18,19} In developing countries, rapid population growth often results in
187 urbanisation, slums, poor sanitation, poverty, subsistence livestock and agricultural intensification –
188 all of which are important drivers of zoonotic disease transmission.^{17,20} The Pacific Islands are
189 particularly vulnerable to the health impacts of climate change because of all of the socio-
190 demographic, geographic, and environmental factors mentioned above,^{21,22} and leptospirosis causes
191 significant health impact in the region.²³⁻²⁸

200 Over the past decades, Fiji has experienced increasing incidence and outbreaks of leptospirosis.^{27,29-}

201
202 ³¹ Two post-flooding outbreaks occurred in 2012, resulting in over 500 cases and 40 deaths. An
203 eco-epidemiological study conducted in 2013 found a community leptospirosis seroprevalence (the
204 percentage of a population with detectable leptospirosis antibodies in their blood) of 19·4% using
205 the microscopic agglutination test (MAT), with significant variation between ethnic groups and
206 residential settings. The findings of the study have been published, focusing on risk factor analysis
207 using standard regression approaches.²⁷ The study provided important insights into leptospirosis
208 eco-epidemiology in Fiji, but there remain multiple unanswered questions with important public
209 health implications. Important questions regarding the reasons for the disparate risk between ethnic
210 groups and residential settings have not been clearly answered, but it is possible that niche-specific
211 interventions may be required for more effective public health control measures. For example,
212 intervention strategies may need a different focus for each ethnic group and/or vary between urban,
213 peri-urban, and rural areas. The study also raised questions about the relative importance of animal
214 species in human infections, a fundamental question when prioritising public health interventions
215 for leptospirosis. On univariate regression analysis, infection was associated with contact with
216 multiple animal species, including rodents, mongoose, dogs, and multiple species of livestock.
217 However, there were significant correlations between presence of different animals species (e.g.
218 people who own pigs are also more likely to own cows), and on multivariable regression analyses,
219 the only animal-related predictor variables retained in the final model were the presence of pigs in
220 the community and high cattle density. Based on these results, can we assume that animal species
221 other than pigs and cattle did not play an important role in human infections? Or could other

237
238
239 species be also important, but excluded from multivariable regression models because they were
240
241 strongly correlated with exposure to pigs or cattle? Also, might the relative importance of different
242
243 animal species differ between ethnic groups and residential settings, and therefore require more
244
245 tailored interventions? These questions highlight some of the limitations of using standard
246
247 regression analysis to model infectious diseases with complex transmission dynamics and
248
249 environmental drivers.

250
251 In this paper, we explore the use of BNs as an alternative methodological approach for modelling
252
253 the eco-epidemiology of infectious diseases, using leptospirosis in Fiji as a case study. Firstly, the
254
255 study aims to improve model performance of BNs by building models that better represent and
256
257 explain causality. Secondly, the study aims to use BNs to determine the relative importance of
258
259 animal species in disease transmission in different ethnic groups and residential settings.

260 261 **MATERIALS and METHODS**

262 263 **Study location and setting**

264
265 Fiji has a population of 837,217³² living in urban, peri-urban, and rural settings in tropical islands.
266
267 Two main ethnic groups, iTaukei (indigenous Fijian) and Indo-Fijians (Fijians of Indian descent),
268
269 account for 57% and 35% of the population respectively.³² Subsistence livestock are common in
270
271 backyards and communal areas, particularly in rural areas. Rodents, mongoose, dogs, and cats are
272
273 abundant in both urban and rural areas.

274 275 **Data sources**

276
277 This study used a database from a recently published study of leptospirosis in Fiji, which was
278
279 designed to include a representative sample of the country's population.²⁷ Briefly, the cross-
280
281 sectional community seroprevalence study included 2,152 participants aged 1 to 90 years from 81
282
283 communities on the three main islands of Fiji. Blood samples were collected from each participant,
284
285 and the microscopic agglutination test (MAT) was used to determine the presence of *Leptospira*
286
287 antibodies, an indicator of previous infection. Data on socio-demographics, environmental factors,
288
289 and animal exposure were obtained from questionnaires, population census, agricultural census,
290
291 World Bank poverty survey, and geo-referenced environmental data. Data were linked to household
292
293 locations using geographic information systems (GIS) to generate a richly structured geospatial
294
295 database that relates risk factors and outcome (presence of *Leptospira* antibodies) for each
individual.

Predictor/indicator variables examined in this study

In this study, we focused on more in-depth analysis of the following predictor/indicator variables, and built scenarios related to animal exposure in different ethnic groups and residential settings:

- Ethnic group:
 - iTaukei, Indo-Fijian (other ethnic groups were excluded because they accounted for only 2% of the study population)
- Residential setting:
 - Urban, peri-urban, rural
- Exposure to animals at the individual/household levels:
 - Physical contact with rodents and/or mongoose
 - Dogs, cats, chickens, pigs, cows, goats, horses
- Exposure to animals at the community level:
 - Pigs, cows, goats, horses

Table 1 provides a summary of the distribution of ethnic groups and residential settings in the study population, and the variations in *Leptospira* seroprevalence found in the 2013 study.

Table 1. Summary of distribution of ethnic groups and residential settings in dataset, and differences in observed seroprevalence in each subgroup.

Variable	Number of subjects	% of total subjects	Observed seroprevalence	Univariate odds ratio (regression analysis)	p value
Total sampled	2152	100%	19.4%		
Ethnic groups					
Indo-Fijian	459	21.3%	7.4%	1	
iTaukei	1651	76.7%	22.7%	3.66	<0.001
Other	39	2.0%	20.5%	3.23	0.114
Residential settings					
Urban	579	26.9%	11.1%	1	
Peri-urban	287	13.3%	15.3%	1.46	0.074
Rural	1286	59.8%	24.0%	2.54	<0.001

Adapted from Lau et al 2016 (27)

The frequency of exposure to animal species in each ethnic group and residential setting were summarised. For individual/household-level analyses, physical contact with rodents or mongoose were included in the analyses but mere sighting of these species around the home were not included

because 85.9% and 77.1% of participants reported sighting of rodents and mongoose respectively; these variables therefore did not provide good discriminatory power and were not statistically associated with the presence of *Leptospira* antibodies at a univariate level. Similarly, the presence of rodents, mongoose, dogs, cats and chickens were not assessed at the community level because these species were ubiquitous.

Bayesian Networks

BNs are probabilistic models based on Bayes' theorem of conditional probability, composed of: i) directed acyclic graphs (DAGs) with nodes that represent variables and outcomes and arrows that define dependency between nodes, and ii) node probability tables (NPT).³³ BNs were constructed using Netica software.³⁴ Figure 1 shows a simple BN, where 'presence of *Leptospira* antibodies' (child node) is dependent on 'pigs in community' and 'residential setting' (parent nodes). 'Pigs in community' is in turn dependent on 'residential setting'. For child nodes that conditionally depend on their parent nodes, the NPT is called a conditional probability table (CPT) that defines the probabilistic relationship between the nodes. The CPT for 'Presence of *Leptospira* antibodies' (Table 2) shows that for a rural setting with pigs, there is a 27.5% probability of the presence of antibodies. For parentless nodes, e.g. 'residential setting', an unconditional probability table stores the prior probabilities of each state: e.g. Figure 1a shows that 59.8% of the population live in rural areas.

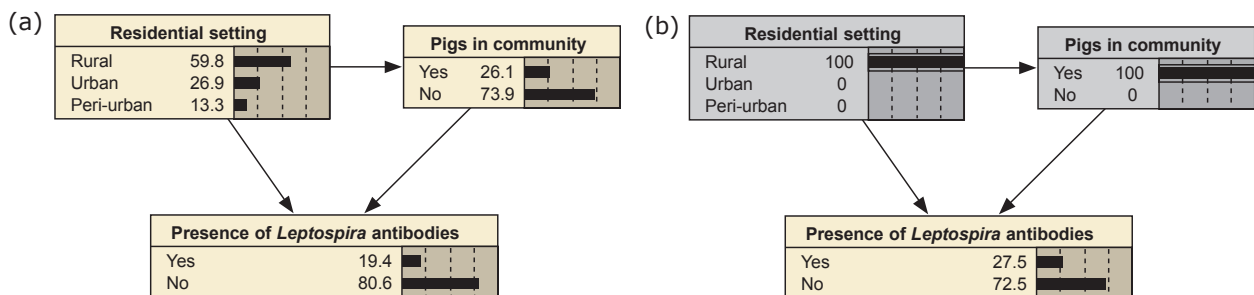


Figure 1. A simple Bayesian network for estimating the probability of the 'Presence of *Leptospira* antibodies' based on the presence/absence of pigs in the community and type of residential setting. The network has two predictor or "parent" nodes ('Pigs in community' and 'Residential setting') linked to the outcome or child node ('Presence of *Leptospira* antibodies'). The presence/absence of 'Pigs in community' is also dependent on 'Residential setting'. The 'Pigs in community' node includes two categories or 'states': Yes or No. The 'Residential setting' variable includes three states: Rural, Urban, and Peri-urban. In Figure 1a), the nodes were set to show the 'default probabilities' in the belief bars, which provide a reflection of the data, i.e. approximately 26.1% of the study population had pigs in their community, 59.8% lived in rural areas, and *Leptospira* antibodies were present in 19.4%. In Figure 1b), a scenario was defined by selecting belief bars to show that in a rural residential setting where pigs were present, the probability of *Leptospira* antibodies being present was 27.5%.

Table 2. Conditional probabilities table (CPT) for the ‘Presence of *Leptospira* antibodies’ node, showing the probabilities of the presence/absence of *Leptospira* antibodies for all combinations of states in the parent nodes (‘Residential setting’ and ‘Pigs in community’)

States of parent nodes		Probability of the Presence of <i>Leptospira</i> antibodies (%)	
Residential setting	Pigs in community	Yes	No
Rural	Yes	27.5	72.5
Rural	No	22.3	77.7
Urban	Yes	23.8	76.2
Urban	No	8.9	91.1
Peri-urban	Yes	25.9	24.1
Peri-urban	No	12.9	87.1

In naïve BNs, predictor/indicator variables are assumed to be independent. In structured BNs, causal dependencies between nodes can be defined using arrows, and each node can be used as predictor or indicator depending on the direction of the arrow. The graphical interface of BNs allows users to define scenarios by selecting states for each node (e.g. a rural community with pigs). When a node state is selected (referred to as inserting findings or evidence), the probabilities in all other nodes are updated using Bayes’ Theorem of conditional probabilities according to the causal dependencies among nodes (probability propagation). NPTs and causal dependency can be learnt directly from data via parameter and structural learning algorithms, or derived from expert opinion.

Model structure and parameterisation

Three groups of BNs, one naïve and two expert-structured, were built and used to analyse scenarios of animal exposure for the two ethnic groups (iTaukei and Indo-Fijian) and three residential settings (urban, peri-urban, rural). Group A BNs were naïve networks, which assumed that all predictor/indicator variables were independent. Group B and C BNs were structured networks designed specifically to examine the role of each animal species in disease transmission in different ethnic groups and residential settings. BNs in Groups A, B, and C were compiled based on the influence diagrams in Figure 2. Table 3 shows the codes of the three groups of BNs for ease of reference.

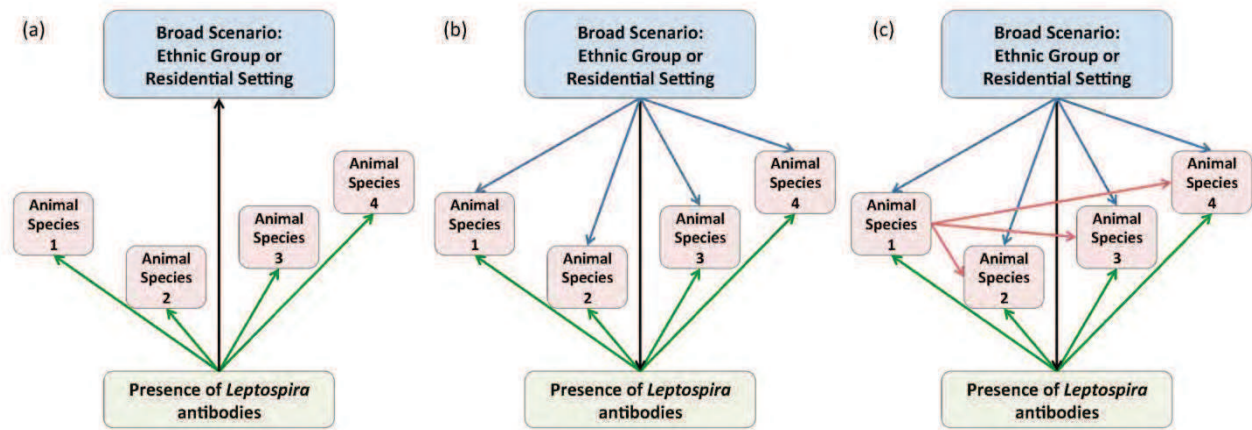


Figure 2. Frameworks for influence diagrams for a) Group A BNs were naïve networks and assume that all indicator variables were independent, with each variable individually linked to the outcome; b) Group B BNs were structured networks, and reflect that the broad scenario is a *predictor* (parent node) of the presence of each animal species (blue arrows), and each animal species is in turn an *indicator* (child node) of the outcome (green arrows); c) Group C were structured to also take into account interdependence between nodes related to animal exposure by creating links from species A to species B, C and D (red arrows). The broad scenario was also directly linked to the outcome (black arrow) to take into account the alternate exposure pathways (other than animal exposure) through which ethnicity and residential setting could influence infection risk (e.g. behaviour, occupation).

The influence diagram for Group A BNs (Figure 2a) assumes that all indicator variables were independent, and each variable was individually linked to the outcome (presence of *Leptospira* antibodies). The influence diagram for Group B BNs (Figure 2b) was structured to reflect that the broad scenario (ethnic group or residential status) is a *predictor* (parent node) of the presence of each animal species in the community (blue arrows), and each animal species is in turn an *indicator* (child node) of the presence of *Leptospira* antibodies (green arrows). Animal species nodes were not used as predictors of the outcome because this structure would have resulted in a very large conditional probability table for the outcome node, and undefined probabilities for a significant number of scenarios. It is more logical to have arrows pointing from cause to effect, but in some cases, the directions of arrows are reversed to avoid large conditional probability tables that are difficult to parameterise with available data. Reversing the direction of arrows is possible in a BN because inference can work both directions.³⁵ However, biological plausibility needs to be considered when determining the direction of causation, which is not necessarily the same as the direction of the arrows. For example, in our models, exposure to animals ‘causes’ an increased risk of leptospirosis, and not vice versa.

BNs in Group C (Figure 2c) were structured to also take into account dependence between the variables related to animal exposure. Links were created between the most common animal species and all other species (red arrows), resulting in conditional probabilities that take into account

532
533
534 dependence between the animal variables, e.g. the presence of cows is correlated with the presence
535 of goats, pigs, and horses. For BNs related to individual/household-level animal exposure, animal
536 species were categorised into three groups: feral (rodents and mongoose), pets (dogs and cats) and
537 livestock (goats, pigs, horses and cows). Dependencies were modelled only within each of the three
538 animal groups. The broad scenario node was also directly linked to the outcome node (black arrow)
539 to take into account the alternate exposure pathways (other than animal exposure) through which
540 ethnicity or residential setting could influence infection risk (for example behaviour, occupation,
541 poverty or sanitation).
542
543

544
545
546
547
548 Conceptually, Group A BNs are similar to standard regression models, where all predictor/indicator
549 variables are independent. Group B BNs were structured to provide a better representation of the
550 causal relationships between variables. Group C also considered interdependence between the
551 animal variables. Unlike standard regression models, BNs are capable of incorporating and
552 retaining strongly correlated variables in the final models, such as exposure to multiple animal
553 species.
554
555
556
557
558
559
560

561 **Model training and testing**

562 Bayesian networks are driven by the Bayes theorem of conditional probability and allows prior
563 knowledge to be incorporated into model predictions. Bayes theorem (Equation 1) states that the
564 conditional probability of a hypothesis (H) occurring given evidence (E), can be calculated as the
565 product of the probability of H and the conditional probability of E given H, divided by the
566 probability of E.
567
568
569

$$570 \quad P(H | E) = P(H) \times P(E | H) / P(E) \quad \text{Equation 1}$$

571
572
573 In a BN, this formula is used to calculate and update conditional probabilities of all node states
574 when evidence is inserted into one or more nodes. Probabilities for NPTs (including CPTs) can be
575 either learnt from the data during model training, or defined by experts.
576
577
578

579 Networks were trained using the Expectation Maximisation algorithm³⁶ in Netica, and tested using
580 two methods:
581

582
583 1. Model discrimination ability was measured using the area under the curve of the receiver
584 operating characteristic (AUC). The AUC for each BN was calculated using trials, where 50% of
585 the data were used to train the BN and populate the CPTs, and the other 50% used to test the BN (to
586
587
588
589
590

591
592
593 determine the accuracy of the predicted prevalence values). For each BN, repeated random
594 subsampling was used to conduct 30 trials, and the average AUC reported.
595
596

597
598 2. Model calibration (measure of how well the model fits the data, or model goodness-of-fit) was
599 measured by comparing predicted and observed probabilities for each set of BNs. For this purpose,
600 BNs were trained using 100% of the dataset. The agreement between predicted probabilities of the
601 presence of *Leptospira* antibodies under different scenarios and the observed probabilities
602 (empirical data from the 2013 field study) were measured using R^2 and mean squared error (MSE).
603
604 We examined scenarios based on ethnicity, residential location, and exposure to animal species.
605
606 After defining a broad scenario of ethnicity or residential location, more specific scenarios of
607 animal exposure were examined. We analysed the influence of each animal species individually,
608 and also combinations of two and three animal species if these scenarios were reported by >3% of
609 at least one ethnic or residential subgroup. Less common scenarios were not assessed because of
610 insufficient data for robust predictions, and their low relevance for understanding disease
611 transmission and informing public health interventions. Nodes that were not included in a scenario
612 were left in their default state. Each trio of Group A, B, and C BNs were compared to determine
613 whether predictive performance of models improved by structures that better represented causality.
614
615
616
617
618
619

620 **Relative importance of animal species under different scenarios**

621
622 The relative importance of each animal species in leptospirosis transmission for each ethnic group
623 and residential setting were examined using the Group C BNs. To ascertain whether exposure to
624 one or more animal species had a significant effect on seroprevalence, a test of proportions was
625 used to determine if differences in predicted seroprevalence between exposed and unexposed
626 groups were statistically significant at $p < 0.05$.
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649

650
651
652 **RESULTS**
653

654
655 **Bayesian network models**
656

657 Based on the influence diagrams in Figure 2, 12 BNs were compiled. Differences between the BNs
658 are summarised in Table 3, and each of the BNs were assigned a code for ease of reference. The
659 structures and variables included in each set of BNs are shown in Figure 3A to 3D. The ‘belief bars’
660 in the figures show the probability distributions for the states of each node captured by the dataset,
661 and reflect conditional probabilities between all connected nodes, e.g. Figure 3B shows that 76.8%
662 of the study population are of iTaukei ethnicity, 26.1% reported the presence of pigs in the
663 community, and 19.4% were seropositive for leptospirosis.
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708

Table 3. Summary of the three groups of BNs used to examine the role of animal species in different ethnic groups and residential settings, and the codes used for each BN for ease of reference.

	Group A	Group B	Group C
Influence diagram	Figure 2a	Figure 2b	Figure 2c
Model type	Naïve Bayesian network	Structured Bayesian network	Structured Bayesian network
Assumptions about predictor/indicator variables	All predictor/indicator variables independent	Variables related to animal exposure were independent, e.g. presence of cows was not correlated with presence of other animal species.	Considered dependence between variables related to animal exposure, e.g. presence of cows was associated with the presence of other animal species.
Model structure	Each predictor/indicator variable individually linked to the outcome. Conceptually similar to regression models.	The broad scenario (ethnic group or residential status) was used as a <i>predictor</i> (parent node) of the presence of each animal species (blue arrows), and each animal species was in turn used as an <i>indicator</i> (child node) of the presence of <i>Leptospira</i> antibodies (green arrows). The broad scenario also directly linked to the outcome node (black arrow) to take into account the alternate exposure pathways (other than animal exposure) through which ethnicity or residential setting could influence infection risk (for example behaviour, occupation, poverty or sanitation).	In addition to the model structures for Group B BNs, Group C BNs also considered dependence between the variables related to animal exposure. Links were created between the most common animal species and all other species (red arrows), resulting in conditional probabilities that take into account dependence between animal variables, e.g. the presence of cows is correlated with the presence of other animal species. For BNs related to individual/household-level animal exposure, animal species were categorised into three groups: feral (rodents and mongoose), pets (dogs and cats) and livestock (goats, pigs, horses and cows). Dependencies were modelled only within each of the three animal groups.
Codes for BNs used to examine Ethnicity and Individual/household-level animal exposure (Figure 3A)	EI-A	EI-B	EI-C
Codes for BNs used to examine Ethnicity and Community-level animal exposure (Figure 3B)	EC-A	EC-B	EC-C
Codes for BNs used to examine Residential setting and Individual/household-level animal exposure (Figure 3C)	RI-A	RI-B	RI-C
Codes for BNs used to examine Residential setting and Community-level animal exposure (Figure 3D)	RC-A	RC-B	RC-C

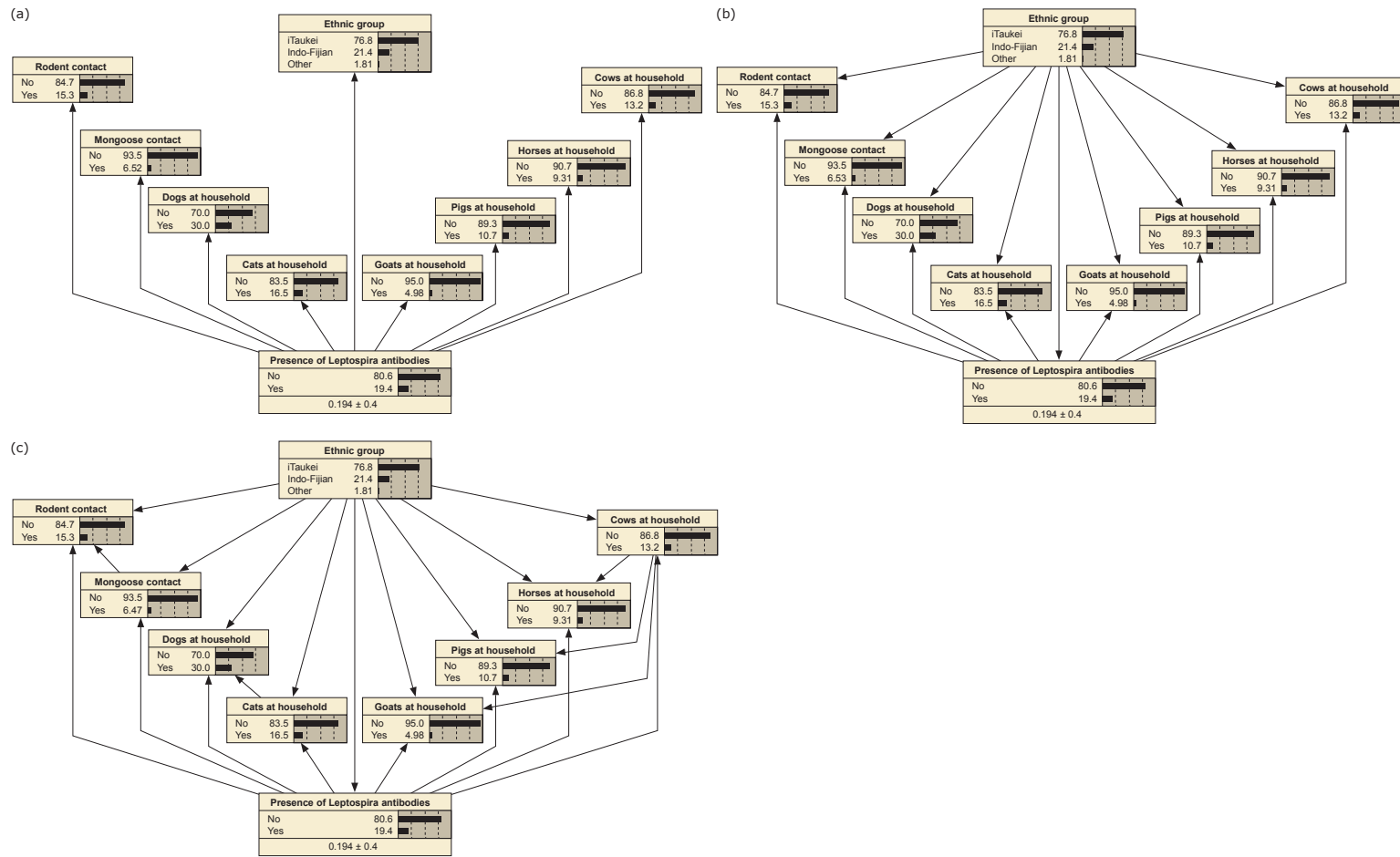


Figure 3A. BNs used to model the probability of the presence of *Leptospira* antibodies based on ethnicity and individual/household-level exposure to livestock animal species. a) BN EI-A, a naïve network assuming that all variables were independent, b) BN EI-B, a structured network that provides a better representation of interrelationships between variables, but assuming that animal variables were independent, and c) BN EI-C, structured network taking into account interdependence between animal variables.

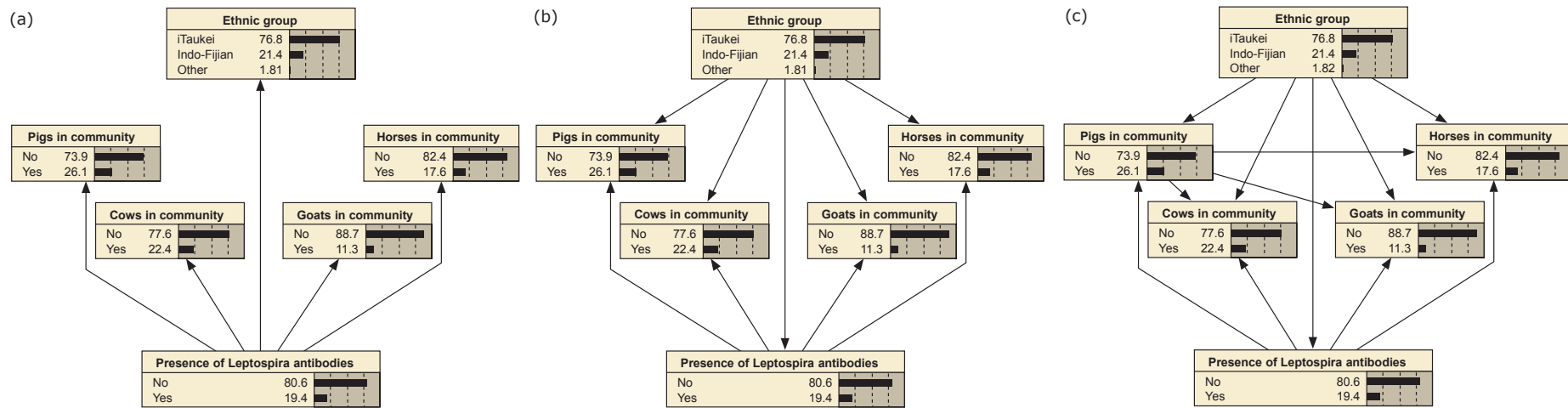


Figure 3B. BNs used to model the probability of the presence of *Leptospira* antibodies based on ethnicity and the presence of livestock animal species in the community: a) BN EC-A, a naïve network assuming that all variables were independent, b) BN EC-B, a structured network that provides a better representation of interrelationships between variables, but assuming that animal variables were independent, and c) BN EC-C, structured network taking into account interdependence between animal variables.

832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872

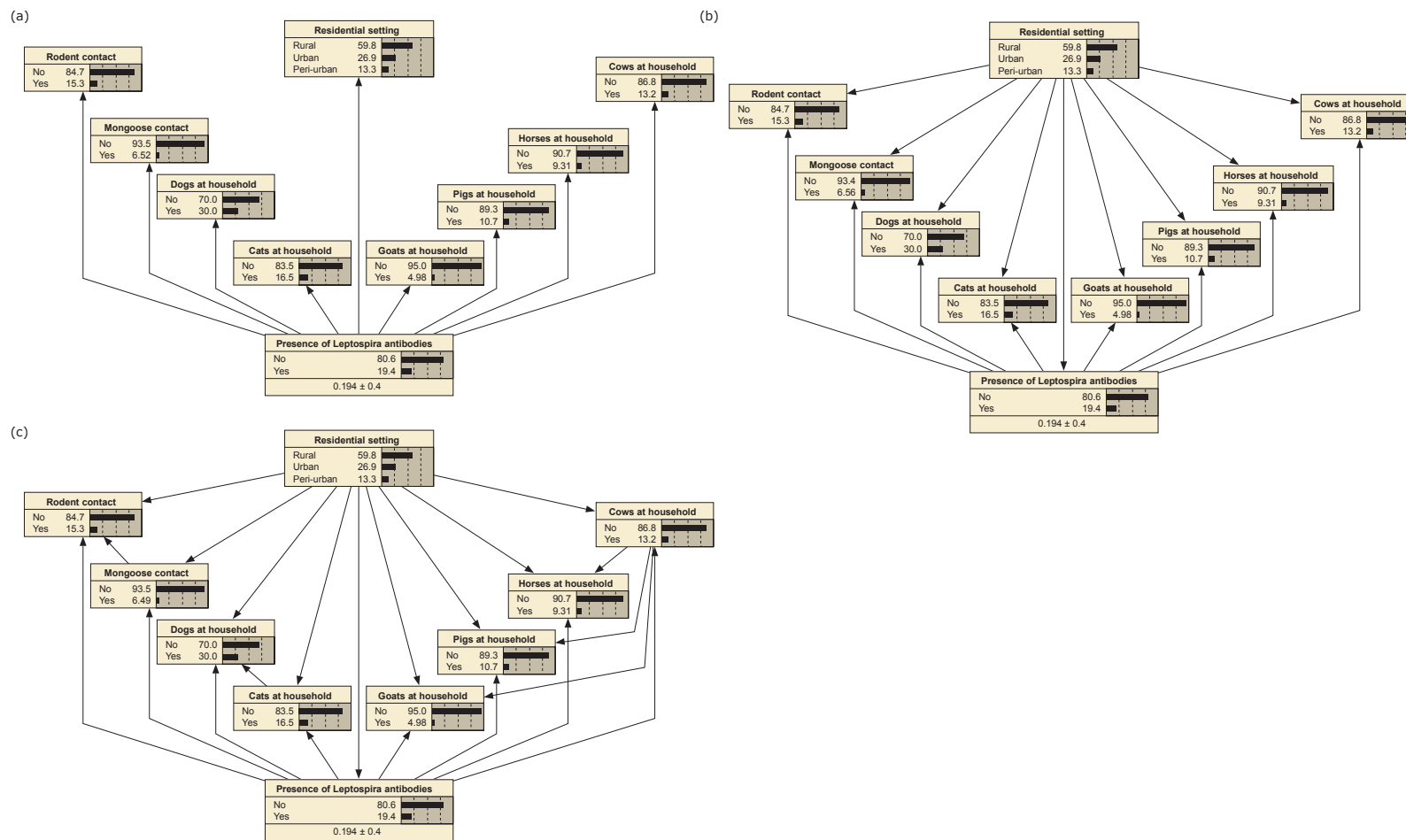


Figure 3C. BNs used to model the probability of the presence of *Leptospira* antibodies based on residential setting and individual/household level exposure to livestock animal species. a) BN RI-A, a naïve network assuming that all variables were independent, b) BN RI-B, a structured network that provides a better representation of interrelationships between variables, but assuming that animal variables were independent, and c) BN RI-C, structured network taking into account interdependence between animal variables.

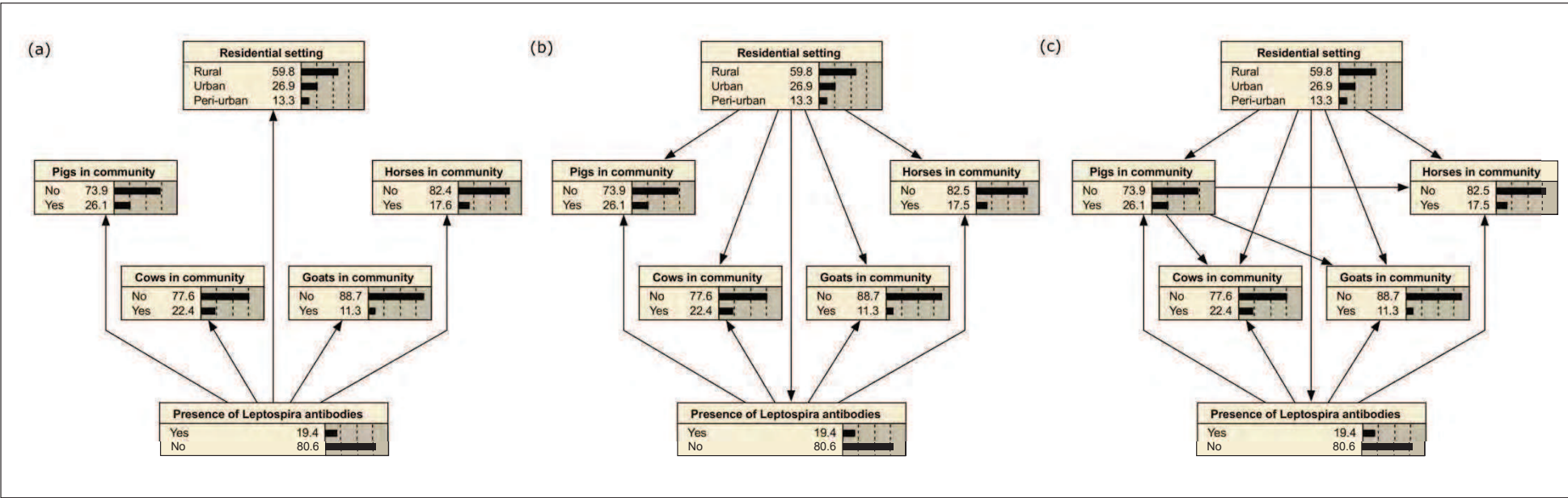


Figure 3D. BNs used to model the probability of the presence of *Leptospira* antibodies based on residential setting and the presence of livestock animal species in the community. a) BN RC-A, a naïve network assuming that all variables were independent, b) BN RC-B, a structured network that provides a better representation of interrelationships between variables, but assuming that animal variables were independent, and c) BN RC-C, structured network taking into account interdependence between animal variables.

914
915
916 **Model testing**
917

918 **a) Model discrimination ability – AUC**
919

920 The median AUC results over the 30 trials for each of the 12 BNs ranged from 0.59-0.61 (Table 4),
921 and indicate poor (but better than random) model discriminatory ability. There were no significant
922 differences in AUCs between Groups A, B and C BNs.
923
924

925 **Table 4. AUC results over 30 trials for Group A, B, and C BNs.**
926

927

Bayesian Network Code	Median AUC	Interquartile Range
Ethnicity and Individual/household-level exposure to animals:		
EI-A	0.61	0.60-0.61
EI-B	0.60	0.59-0.61
EI-C	0.59	0.58-0.60
Ethnicity and Community-level exposure to animals:		
EC-A	0.61	0.60-0.62
EC-B	0.61	0.60-0.62
EC-C	0.60	0.61-0.63
Residential setting and Individual/household-level exposure to animals:		
RI-A	0.61	0.60-0.61
RI-B	0.60	0.58-0.60
RI-C	0.59	0.58-0.60
Residential setting and Community-level exposure to animals:		
RC-A	0.61	0.61-0.62
RC-B	0.60	0.60-0.61
RC-C	0.60	0.59-0.61

928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953

954
955
956
957 **b) Model calibration – predicted versus observed seroprevalence**
958

959 Tables 5 to 8 show the scenarios of animal exposure for ethnic group and residential setting where
960 at least 3% of one or more subgroups reported the exposure scenarios; these scenarios were
961 included in further analyses. The tables also show the percentage of each subgroup that reported the
962 animal exposures. For example, Table 6 shows the most common scenarios of community-level
963 animal exposure(s) for each ethnic group, where at least 3% of one or more ethnic group reported
964 that combination of animal exposure. Sections A, B, and C list the scenarios related to exposure to
965
966
967
968
969

973
974
975 each animal species, combinations of two animal species, and combinations of three animal species
976 respectively. If a scenario was reported by <3% of a subgroup, the predicted seroprevalence is not
977 reported.
978
979

980
981 For each scenario of animal exposure shown in Tables 5 to 8, the predicted seroprevalence were
982 calculated using the associated BNs and compared to the observed seroprevalence. For example,
983 BNs EC-A, EC-B, and EC-C were used to predict seroprevalence for each of the scenarios of
984 ethnicity and community-level animal exposure(s) shown in Table 6. Section B of Table 6 shows
985 that 16.7% of iTaukei and 4.4% of Indo-Fijians reported the presence of both cows *and* horses in
986 their community. And in iTaukei who reported exposure to both cows and horses, the observed
987 seroprevalence was 25.5%, while the predicted seroprevalence using EC-A, EC-B, and EC-C were
988 36.3%, 29.4%, and 27.3% respectively.
989
990

991
992 Agreement between predicted and observed seroprevalence were measured using R^2 and MSE, and
993 the correlations for each trio of Group A, B, and C models are shown in Figures 4 and 5. The
994 figures show that R^2 values improved from 0.59 for EI-A to 0.93 for EI-C; 0.78 for EC-A to 0.93
995 for EC-C; 0.54 for RI-A to 1.00 for RI-C; and 0 for RC-A to 0.75 for RC-C. Similarly, MSE
996 showed that Group C models produced the best agreement between predicted and observed
997 seroprevalence. MSE were 67.1, 22.6 and 3.6 for EI-A, EI-B, and EI-C; 95.0, 67.2, and 7.1 for EC-
998 A, EC-B, and EC-C; 46.8, 6.3, and 0.3 for RI-A, RI-B, and RI-C; and 144.8, 364.3, and 16.6 for
999 RC-A, RC-B, and RC-C respectively. For each trio of BNs, the best predictive accuracy (highest R^2
1000 and lowest MSE) was seen with Group C models.
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031

1032
1033
1034
1035
1036
1037
1038
1039
1040
1041
1042
1043
1044
1045
1046
1047
1048
1049
1050
1051
1052
1053
1054
1055
1056
1057
1058
1059
1060
1061
1062
1063
1064
1065
1066
1067
1068
1069
1070
1071
1072

Table 5. The most common individual/household-level exposure to animal species in each ethnic group. For rodents and mongoose, exposure was defined as physical contact with these animals. For other animal species, exposure was defined as presence of the animal species at the individual’s household. BNs EI-A, EI-B and EI-C were used to predict seroprevalence under each of the scenarios shown below, and summarised in Figure 4a.

Section	Physical contact		Animal species present at household						% of population exposed to animal species		Observed seroprevalence* (%)		Predicted seroprevalence using EI-A (%)		Predicted seroprevalence using EI-B (%)		Predicted seroprevalence using EI-C (%)	
	Rodents	Mongoose	Dog	Cat	Cow	Goat	Horse	Pig	iTaukei n=1651	Indo-Fijian n=459	iTaukei	Indo-Fijian	iTaukei	Indo-Fijian	iTaukei	Indo-Fijian	iTaukei	Indo-Fijian
A. Exposure scenarios related to EACH animal species	X								17.3	6.3	27.3	6.9	30.0	10.5	27.8	7.1	27.9	7.1
		X							7.5	2.0	30.1	0.0	33.7	-	30.5	-	30.6	-
			X						26.0	43.6	24.0	9.5	22.7	7.42	24.1	9.5	24.1	9.5
				X					14.4	23.3	19.4	9.4	19.3	6.11	19.4	9.3	19.4	9.3
					X				14.1	10.7	27.0	18.4	29.7	10.3	27.1	18.4	27.1	18.4
						X			3.1	12.2	23.5	17.9	24.0	7.95	23.5	17.8	23.5	17.8
							X		10.5	5.7	27.0	19.2	30.0	10.5	27.0	19.2	27.0	19.2
								X	13.7	0.4	25.7	-	30.1	-	25.7	-	25.7	-
B. Exposure scenarios related to combinations of TWO animal species	X	X							4.6	1.1	29.0	0.0	42.6	-	36.6	-	29.7	-
			X	X					6.8	17.2	19.5	11.4	19.3	6.12	20.7	11.9	19.5	11.4
			X			X			1.5	8.1	24.0	21.6	-	7.95	-	22.2	-	22.2
			X		X				6.2	7.8	31.1	19.4	29.7	10.4	28.7	22.8	28.7	22.8
C. Exposure scenarios related to combinations of THREE animal species			X	X		X			0.7	4.8	8.3	18.2	-	6.56	-	26.9	-	25.8
			X	X	X				1.8	3.9	13.8	27.8	-	8.58	-	27.5	-	26.5
			X		X	X			1.3	4.4	18.2	25.0	-	11.1	-	44.5	-	28.2
			X		X		X		3.5	2.6	29.3	16.7	38.2	-	33.7	-	30.3	-

*Overall observed seroprevalence in 2013 field study was 22.7% in iTaukei and 7.4% in Indo-Fijians. Predicted seroprevalence were only calculated for animal exposure scenarios reported by >3% of at least one subgroup; “-“ indicates scenarios where predicted seroprevalence were not calculated.

1073
 1074
 1075
 1076
 1077
 1078
 1079
 1080
 1081
 1082
 1083
 1084
 1085
 1086
 1087
 1088
 1089
 1090
 1091
 1092
 1093
 1094
 1095
 1096
 1097
 1098
 1099
 1100
 1101
 1102
 1103
 1104
 1105
 1106
 1107
 1108
 1109
 1110
 1111
 1112
 1113

Table 6. The most common community-level exposure to animal species in each ethnic group. Exposure was defined as the presence of the animal species at the individual's community. BNs EC-A, EC-B and EC-C were used to predict seroprevalence under each of the scenarios shown below, and summarised in Figure 4b.

Section	Animal species present in community				% of population exposed to animal species		Observed seroprevalence* (%)		Predicted seroprevalence using EC-A (%)		Predicted seroprevalence using EC-B (%)		Predicted seroprevalence using EC-C (%)	
	Cow	Goat	Horse	Pig	iTaukei n=1651	Indo-Fijian n=459	iTaukei	Indo-Fijian	iTaukei	Indo-Fijian	iTaukei	Indo-Fijian	iTaukei	Indo-Fijian
A. Exposure scenarios related to EACH animal species	X				24.8	12.6	25.6	12.1	28.6	9.88	25.6	12.1	25.6	12.1
		X			11.3	11.3	28.9	11.5	29.2	10.1	28.9	11.5	28.9	11.5
			X		21.1	5.4	26.2	16.0	29.4	10.2	26.2	16.0	26.2	16.0
				X	32.7	1.3	25.9	16.7	30.8	-	26.1	-	26.1	-
B. Exposure scenarios related to combinations of TWO animal species	X	X			10.7	6.8	30.7	9.7	36.0	13.3	32.3	18.3	29.7	14.0
	X		X		16.7	4.4	25.5	10.0	36.3	13.5	29.4	24.6	27.3	18.5
	X			X	18.6	0.9	26.4	0.0	37.9	-	29.3	-	26.4	-
		X	X		9.6	4.1	29.1	15.8	36.9	13.8	32.9	23.7	30.3	17.8
		X		X	9.5	0.7	29.3	33.3	38.5	-	32.8	-	29.3	-
			X	X	15.9	0.4	27.0	0.0	38.8	-	29.9	-	27.0	-
C. Exposure scenarios related to combinations of THREE animal species	X	X	X		9.4	3.5	29.5	6.3	44.5	18.0	36.6	34.7	30.7	13.0
	X	X		X	9.1	0.4	30.5	0.0	46.1	-	36.5	-	29.7	-
	X		X	X	13.4	0.4	26.2	0.0	46.5	-	33.4	-	27.3	-
		X	X	X	8.4	0.4	29.7	0.0	47.1	-	37.1	-	30.3	-

*Overall observed seroprevalence in 2013 field study was 22.7% in iTaukei and 7.4% in Indo-Fijians. Predicted seroprevalence were only calculated for animal exposure scenarios reported by >3% of at least one subgroup; “-” indicates scenarios where predicted seroprevalence were not calculated.

Table 7. The most common individual/household-level exposure to animal species in each residential setting. For rodents and mongoose, exposure was defined as physical contact with these animals. For other animal species, exposure was defined as presence of the animal species at the individual’s household. BNs RI-A, RI-B and RI-C were used to predict seroprevalence under each of the scenarios shown below, and summarised in Figure 5a.

Section	Physical contact		Animal species present at household						% of population exposed to animal species			Observed seroprevalence* (%)			Predicted seroprevalence using RI-A (%)			Predicted seroprevalence using RI-B (%)			Predicted seroprevalence using RI-C (%)		
	Rodents	Mongoose	Dog	Cat	Cow	Goat	Horse	Pig	Urban n=579	Peri-urban n=287	Rural n=1286	Urban	Peri-urban	Rural	Urban	Peri-urban	Rural	Urban	Peri-urban	Rural	Urban	Peri-urban	Rural
A. Exposure scenarios related to EACH animal species	X								13.6	12.9	16.1	16.5	27.0	29.0	15.4	20.9	31.6	16.3	26.9	29.5	16.3	26.9	29.5
		X							4.0	4.2	7.8	17.4	25.0	32.0	17.7	23.9	35.4	17.1	25.4	32.6	17.2	24.9	32.6
			X						26.6	33.4	30.7	12.3	13.5	23.5	11.1	15.3	24.0	12.3	13.5	23.6	12.3	13.5	23.6
				X					15.2	22.3	15.8	5.7	14.1	21.7	9.2	12.8	20.5	5.7	14.0	21.7	5.7	14.0	21.7
					X				3.5	5.2	19.4	25.0	6.7	26.9	15.2	20.7	31.3	25.0	6.6	27.0	25.0	6.6	27.0
						X			0.9	1.0	7.7	22.0	0.0	21.2	-	-	25.4	-	-	21.2	-	-	21.2
							X		1.4	2.1	14.5	0.0	16.7	27.4	-	-	31.6	-	-	27.5	-	-	27.5
B. Exposure scenarios related to combinations of TWO animal species	X	X							2.8	3.1	4.6	18.8	33.3	30.5	-	31.5	44.5	-	40.8	39.0	-	33.2	31.4
			X	X					5.4	14.3	9.8	3.2	17.1	19.1	9.2	12.8	20.5	6.4	12.3	21.3	3.2	17.0	19.1
			X		X				0.7	2.4	10.0	50.0	0.0	28.9	-	-	31.3	-	-	26.5	-	-	26.5
					X		X		0.7	1.7	9.5	0.0	0.0	29.5	-	-	40.0	-	-	30.6	-	-	29.6
C. Exposure scenarios related to combinations of THREE animal species			X		X		X	0.2	1.0	5.1	0.0	-	28.8	-	-	40.0	-	-	30.1	-	-	29.0	

*Overall observed seroprevalence in 2013 field study was 11.1% in urban, 15.3% in peri-urban, and 24.0% in rural areas. Predicted seroprevalence were only calculated for animal exposure scenarios reported by >3% of at least one subgroup; “-“ indicates scenarios where predicted seroprevalence were not calculated.

Table 8. The most common community-level exposure to animal species in each residential setting. Exposure was defined as the presence of the animal species at the individual's community. BNs RC-A, RC-B and RC-C were used to predict seroprevalence under each of the scenarios shown below, and summarised in Figure 5b.

Section	Animal species present in community				% of population exposed to animal species			Observed seroprevalence* (%)			Predicted seroprevalence using RC-A (%)			Predicted seroprevalence using RC-B (%)			Predicted seroprevalence using RC-C (%)		
	Cow	Goat	Horse	Pig	Urban n=579	Peri-urban n=287	Rural n=1286	Urban	Peri-urban	Rural	Urban	Peri-urban	Rural	Urban	Peri-urban	Rural	Urban	Peri-urban	Rural
A. Exposure scenarios related to EACH animal species	X				9.3	14.3	30.0	24.1	22.0	25.1	14.5	19.9	30.2	24.0	21.9	25.2	24.0	21.9	25.2
		X			7.1	8.0	13.8	29.3	17.4	25.3	14.8	20.3	30.7	29.2	17.3	25.3	29.2	17.3	25.3
			X		8.1	7.0	24.1	25.5	30.0	25.2	15.0	20.5	31.0	25.5	29.9	25.2	25.5	29.9	25.2
				X	14.5	18.8	32.9	23.8	25.9	27.4	15.9	21.6	32.5	23.8	25.8	27.5	23.8	25.8	27.5
B. Exposure scenarios related to combinations of TWO animal species	X	X			6.7	7.3	11.7	30.8	19.1	28.0	19.3	25.8	37.8	51.3	24.5	26.5	35.5	13.8	28.8
	X		X		7.8	6.3	18.4	26.7	22.2	24.2	19.5	26.1	38.1	46.6	39.8	26.4	32.2	23.7	26.3
	X			X	7.6	7.3	20.1	27.3	23.8	27.5	20.5	27.4	39.7	44.3	35.0	28.7	27.2	23.7	27.6
		X	X		6.6	5.2	9.8	31.6	26.7	27.0	19.9	26.5	38.7	53.2	33.1	26.5	37.0	17.9	28.6
		X		X	6.6	3.8	8.9	31.6	9.1	30.7	20.9	27.9	40.3	50.9	28.8	28.9	31.5	9.06	30.7
			X	X	7.3	3.8	16.7	28.6	18.2	27.0	21.2	28.1	40.6	46.2	45.1	28.7	28.5	18.1	27.0
C. Exposure scenarios related to combinations of THREE animal species	X	X	X		6.6	5.2	9.4	31.6	26.7	26.5	25.3	33.1	46.4	74.3	43.3	27.8	41.4	10.2	29.9
	X	X		X	6.6	3.5	8.4	31.6	10.0	31.5	26.6	34.6	48.0	72.5	38.4	30.2	35.6	8.2	30.9
	X		X	X	7.1	3.5	13.6	29.3	10.0	26.3	26.9	34.9	48.3	68.6	56.0	30.0	32.4	16.5	27.1
		X	X	X	6.4	3.1	7.5	32.4	11.1	30.2	27.4	35.4	49.0	74.1	48.8	30.2	37.1	6.0	30.3

*Overall observed seroprevalence in 2013 field study was 11.1% in urban, 15.3% in peri-urban, and 24.0% in rural areas. Predicted seroprevalence were only calculated for animal exposure scenarios reported by >3% of at least one subgroup; “-“ indicates scenarios where predicted seroprevalence were not calculated.

1196
 1197
 1198
 1199
 1200
 1201
 1202
 1203
 1204
 1205
 1206
 1207
 1208
 1209
 1210
 1211
 1212
 1213
 1214
 1215
 1216
 1217
 1218
 1219
 1220
 1221
 1222
 1223
 1224
 1225
 1226
 1227
 1228
 1229
 1230
 1231
 1232
 1233
 1234
 1235
 1236
 1237
 1238
 1239
 1240
 1241
 1242
 1243
 1244
 1245
 1246
 1247
 1248
 1249
 1250
 1251
 1252
 1253
 1254

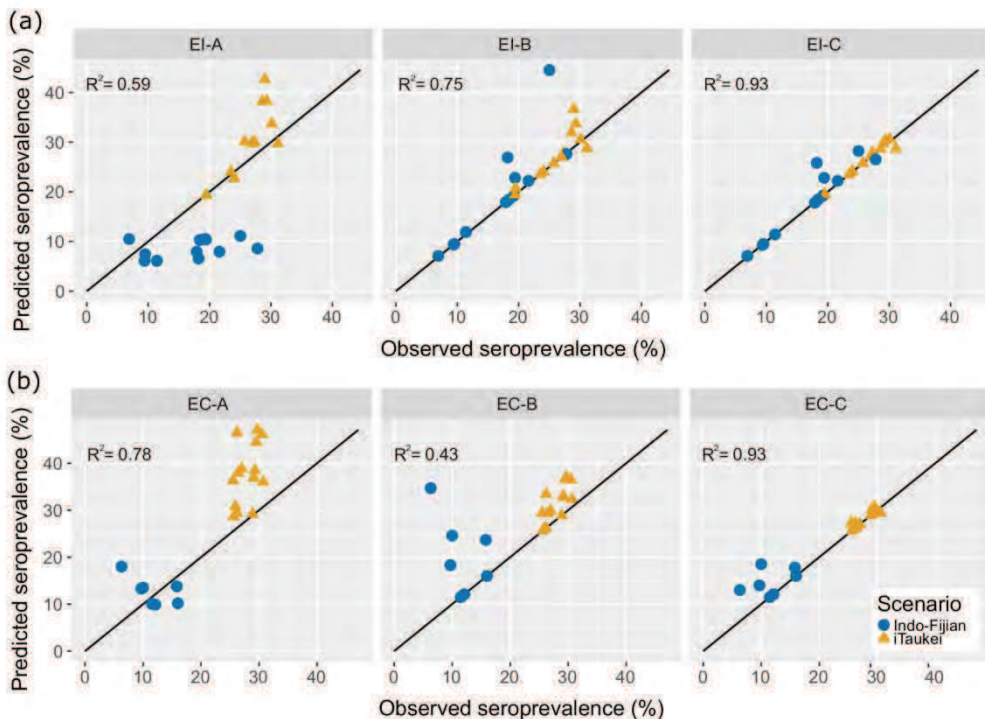


Figure 4. a) Comparison between observed and predicted seroprevalence using Bayesian networks EI-A, EI-B, and EI-C models for individual/household-level exposure for each ethnic group. b) Comparison between observed and predicted seroprevalence using Bayesian networks EC-A, EC-B, and EC-C models for community-level exposure for each ethnic group.

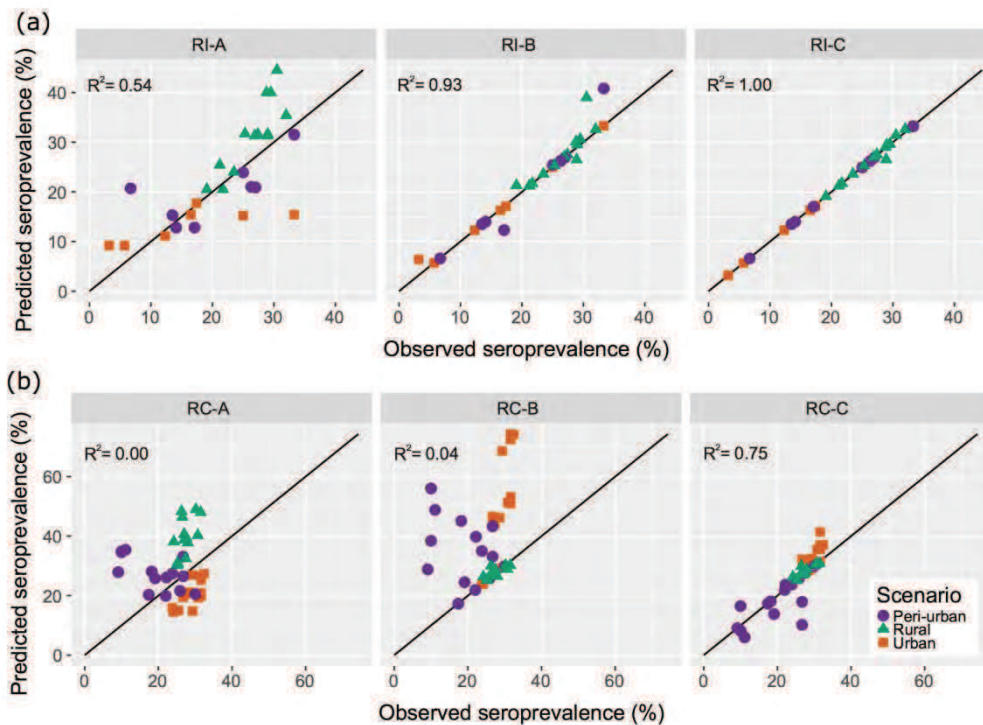


Figure 5. a) Comparison between observed and predicted seroprevalence using Bayesian networks RI-A, RI-B, and RI-C models for individual/household-level exposure and each residential setting. b) Comparison between observed and predicted seroprevalence using Bayesian networks RC-A, RC-B, and RC-C models for community-level exposure and each residential setting.

Relative importance of animal species under different exposure scenarios

Group C BNs showed the best predictive performance, and were used to determine the relative importance of animal species under different scenarios of ethnicity and residential setting. Table 9 shows results of scenario analyses for individual/household-level exposures in ethnic groups (BN EI-C). The prevalence of animal exposures differed markedly between the two ethnic groups, and the animal species associated with higher seroprevalence also varied. For example, 12.2% of Indo-Fijians owned goats, and this scenario was associated with a higher seroprevalence of 17.8% compared to Indo-Fijians who do not own goats (6.0%, $p=0.002$). Only 3.1% of iTaukei owned goats, but this ethnic group was more likely to report physical contact with rodents (17.3%), and this exposure was associated with higher seroprevalence (27.9%) compared to those who do not have contact with rodents (21.6%, $p=0.021$). Figure 6a highlights differences in individual/household animal exposure between ethnic groups, and relative importance of each species on seroprevalence. Triangles and circles represent statistically significant or insignificant differences in seroprevalence between exposed and un-exposed groups.

Table 10 shows the results of scenario analyses for community-level exposures in ethnic groups (BN EC-C). The most common livestock animals found in iTaukei communities were pigs (32.7%) and cows (24.8%). Many communities had multiple livestock species, e.g. 13.4% of iTaukei communities reported the presence of cows *and* pigs *and* horses, and this scenario was associated with a higher predicted seroprevalence of 27.3% compared to communities without any of those animal species (20.6%, $p=0.030$). In contrast, the most common livestock in Indo-Fijian communities were cows (12.6%) and goats (11.3%). Although only 8.7% of Indo-Fijian communities reported the presence of two or more livestock species, the presence of cows *and* horses (reported by 4.4% of Indo-Fijians) was associated with a higher predicted seroprevalence of 18.5% compared to 6.3% in those who were not exposed to these species ($p=0.036$). Figure 6b highlights the differences in exposure and relative importance of animal exposures between ethnic groups.

Table 9. Difference in seroprevalence based on ethnicity and individual/household-level exposure to animal species or combinations of species. BN EI-C was used to predict seroprevalence in exposed and unexposed groups. Results for individual species are summarized in Figure 6a.

Physical contact		Animal species present at household						% of population exposed to animal species		Predicted seroprevalence in exposed (%)		Predicted seroprevalence in unexposed (%)		<i>p</i> value for statistical difference in seroprevalence between exposed and unexposed#	
Rodents	Mongoose	Dog	Cat	Cow	Goat	Horse	Pig	iTaukei n=1651	Indo-Fijian n=459	iTaukei	Indo-Fijian	iTaukei	Indo-Fijian	iTaukei	Indo-Fijian
X								17.3	6.3	27.9	7.1	21.6	7.4	0.021	0.953
	X							7.5	2.0	30.6	-	22.0	-	0.029	-
		X						26.0	43.6	24.1	9.5	22.2	5.8	0.419	0.135
			X					14.4	23.3	19.4	9.3	23.2	6.9	0.196	0.408
				X				14.1	10.7	27.1	18.4	22.0	6.1	0.085	0.002
					X			3.1	12.2	23.5	17.8	22.7	6.0	0.893	0.002
						X		10.5	5.7	27.0	19.2	22.2	6.7	0.153	0.018
							X	13.7	0.4	25.7	-	22.2	-	0.243	-
X	X							4.6	1.1	29.7	-	21.2	-	0.081	-
		X	X					6.8	17.2	19.5	11.4	22.5	6.1	0.465	0.122
		X			X			1.5	8.1	-	22.2	-	4.6	-	<0.001
		X		X				6.2	7.8	28.7	22.8	21.5	4.8	0.092	<0.001
				X		X		7.2	2.6	28.6	-	21.9	-	0.093	-
		X	X		X			0.7	4.8	-	25.8	-	4.9	-	<0.001
		X	X	X				1.8	3.9	-	26.5	-	5.0	-	<0.001
		X		X	X			1.3	4.4	-	28.2	-	4.3	-	<0.001
		X		X		X		3.5	2.6	30.3	-	21.4	-	0.110	-

*Overall observed seroprevalence in 2013 field study was 22.7% in iTaukei and 7.4% in Indo-Fijians. #Using test of difference between proportions, statistically significant results ($p < 0.05$) in bold.

Table 10. Difference in seroprevalence based on ethnicity and community-level exposure to animal species or combinations of species. BN EC-C was used to predict seroprevalence in exposed and unexposed groups. Results for individual species are summarized in Figure 6b.

Animal species present in community				% of population exposed to animal species		Predicted seroprevalence in exposed (%)		Predicted seroprevalence in unexposed (%)		<i>p</i> value for statistical difference in seroprevalence between exposed and unexposed#	
Cow	Goat	Horse	Pig	iTaukei n=1651	Indo-Fijian n=459	iTaukei	Indo-Fijian	iTaukei	Indo-Fijian	iTaukei	Indo-Fijian
X				24.8	12.6	25.6	12.1	21.7	6.7	0.102	0.142
	X			11.3	11.3	28.9	11.5	21.9	6.9	0.031	0.233
		X		21.1	5.4	26.2	16.0	21.7	6.9	0.075	0.091
			X	32.7	1.3	26.1	-	21.0	-	0.020	-
X	X			10.7	6.8	29.7	14.0	21.2	6.2	0.011	0.097
X		X		16.7	4.4	27.3	18.5	21.1	6.3	0.026	0.036
X			X	18.6	0.9	26.4	-	20.8	-	0.038	-
	X	X		9.6	4.1	30.3	17.8	21.2	6.4	0.009	0.056
	X		X	9.5	0.7	29.3	-	20.9	-	0.018	-
		X	X	15.9	0.4	27.0	-	20.8	-	0.031	-
X	X	X		9.4	3.5	30.7	13.0	20.7	5.7	0.005	0.228
X	X		X	9.1	0.4	29.7	-	20.6	-	0.012	-
X		X	X	13.4	0.4	27.3	-	20.6	-	0.030	-
	X	X	X	8.4	0.4	30.3	-	20.7	-	0.010	-

*Overall observed seroprevalence in 2013 field study was 22.7% in iTaukei and 7.4% in Indo-Fijians.

#Using test of difference between proportions, statistically significant results ($p < 0.05$) in bold.

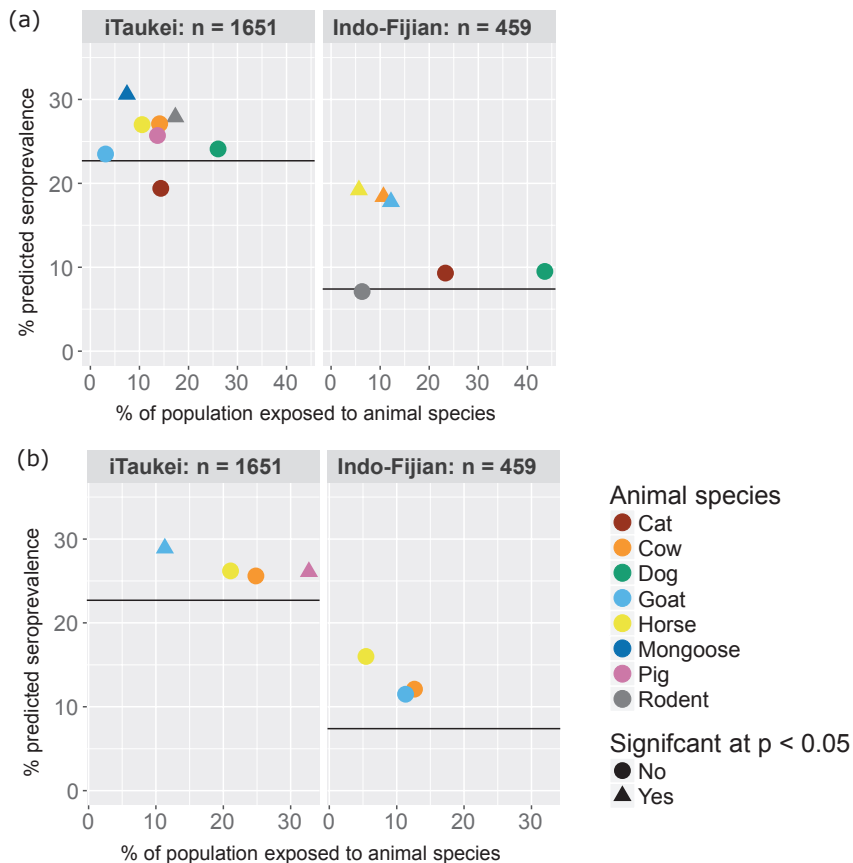


Figure 6. a) Individual/household-level exposure to animals – differences in exposure and predicted seroprevalence between ethnic groups. Exposure is defined as physical contact with rodents or mongoose, or presence of other animal species at the individual’s household. b) Community-level exposure to animals – differences in exposure and predicted seroprevalence between ethnic groups. Exposure is defined as the presence of animal species at the individual’s community. Horizontal black lines indicate mean seroprevalence for each subgroup. Triangles/circles indicate statistically significant/insignificant difference in seroprevalence between exposed and un-exposed groups.

Table 11 shows the results of scenario analyses for individual/household-level exposures in different residential settings (BN RI-C). In urban areas, the most common animal exposures were to dogs (26.6%), cats (15.2%), and rodents (13.6%). Few urban residents reported exposure to cows (3.5%) or pigs (3.6%), but their presence at households was associated with a higher predicted seroprevalences of 25.0% (vs 10.6%, $p=0.044$) and 33.3% (vs 10.2%, $p<0.001$) compared to those without these exposures. In rural areas, physical contact with rodents (16.1%) and mongoose (7.8%) were more common than in urban or peri-urban areas, and associated with higher seroprevalence of 29.5% (vs 22.9%, $p=0.042$) and 32.6% (vs 23.3%, $p=0.037$). Figure 7a highlights the differences in exposure and relative importance of individual/household-level animal exposures between urban, peri-urban, and rural areas.

1455
1456
1457
1458
1459 Table 12 provides results of scenario analyses for community-level exposures in residential settings
1460 (BN RC-C). Pigs were the most common livestock species in all community types, present in
1461 14.5% of urban, 18.8% of peri-urban, and 32.9% of rural communities. Pigs were associated with
1462 higher seroprevalence in all community types, but particularly striking in urban areas where
1463 exposure was associated with a seroprevalence of 23.8%, compared to 8.9% in urban dwellers who
1464 were not exposed to pigs ($p<0.001$). Multiple livestock species in urban areas was associated with
1465 very high predicted seroprevalence, e.g. 35.6% in urban communities with cows *and* goats *and* pigs
1466 ($p<0.001$). Figure 7b highlights the relative importance of animal species in each residential setting.
1467
1468
1469
1470
1471
1472
1473
1474
1475
1476
1477
1478
1479
1480
1481
1482
1483
1484
1485
1486
1487
1488
1489
1490
1491
1492
1493
1494
1495
1496
1497
1498
1499
1500
1501
1502
1503
1504
1505
1506
1507
1508
1509
1510
1511
1512
1513

1514
1515
1516
1517 **Table 11. Difference in seroprevalence based on residential setting and individual/household-level exposure to animal species or combinations of species.**
1518 **BN RI-C was used to predict seroprevalence in exposed and unexposed groups. Results for individual species are summarized in Figure 7a.**
1519

Physical contact		Animal species present at household						% of population exposed to animal species			Predicted seroprevalence in exposed (%)			Predicted seroprevalence in unexposed (%)			<i>p</i> value for statistical difference in seroprevalence between exposed and unexposed#		
Rodents	Mongoose	Dog	Cat	Cow	Goat	Horse	Pig	Urban n=579	Peri-urban n=287	Rural n=1286	Urban	Peri-urban	Rural	Urban	Peri-urban	Rural	Urban	Peri-urban	Rural
X								13.6	12.9	16.1	16.3	26.9	29.5	10.2	13.6	22.9	0.108	0.036	0.042
	X							4.0	4.2	7.8	17.2	24.9	32.6	10.8	14.9	23.3	0.338	0.347	0.037
		X						26.6	33.4	30.7	12.3	13.5	23.6	10.6	16.3	24.2	0.564	0.535	0.816
			X					15.2	22.3	15.8	5.7	14.0	21.7	12.0	15.7	24.5	0.083	0.739	0.392
				X				3.5	5.2	19.4	25.0	6.6	27.0	10.6	15.8	23.3	0.044	0.336	0.220
					X			0.9	1.0	7.7	-	-	21.2	-	-	24.3	-	-	0.488
						X		1.4	2.1	14.5	-	-	27.5	-	-	23.4	-	-	0.226
							X	3.6	6.6	14.8	33.3	26.2	25.3	10.2	14.6	23.8	<0.001	0.176	0.655
X	X							2.8	3.1	4.6	-	33.2	31.4	-	0	22.5	-	<0.001	0.115
		X	X					5.4	14.3	9.8	3.2	17.0	19.1	11.2	17.3	24.1	0.164	0.964	0.218
		X		X				0.7	2.4	10.0	-	-	26.5	-	-	23.5	-	-	0.461
				X		X		0.7	1.7	9.5	-	-	29.6	-	-	23.3	-	-	0.125
		X		X		X		0.2	1.0	5.1	-	-	29.0	-	-	23.5	-	-	0.316

1546 *Overall observed seroprevalence in 2013 field study was 11.1% in urban, 15.3% in peri-urban, and 24.0% in rural areas.

1547 #Using test of difference between proportions, statistically significant results ($p < 0.05$) in bold.
1548
1549
1550
1551
1552
1553
1554

1555
 1556
 1557
 1558 **Table 12. Difference in seroprevalence based on residential setting and community-level exposure to animal species or combinations of species.**
 1559 **BN RC-C was used to predict seroprevalence in exposed and unexposed groups. Results for individual species are summarized in Figure 7b.**
 1560

Animal species present in community				% of population exposed to animal species			Predicted seroprevalence in exposed (%)			Predicted seroprevalence in unexposed (%)			<i>p</i> value for statistical difference in seroprevalence between exposed and unexposed#		
Cow	Goat	Horse	Pig	Urban n=579	Peri-urban n=287	Rural n=1286	Urban	Peri-urban	Rural	Urban	Peri-urban	Rural	Urban	Peri-urban	Rural
X				9.3	14.3	30.0	24.0	21.9	25.2	9.7	14.2	23.5	0.001	0.205	0.513
	X			7.1	8.0	13.8	29.2	17.3	25.3	9.7	15.2	23.8	0.001	0.789	0.664
		X		8.1	7.0	24.1	25.5	29.9	25.2	9.8	14.2	23.7	0.001	0.060	0.590
			X	14.5	18.8	32.9	23.8	25.8	27.5	8.9	12.9	22.3	<0.001	0.018	0.040
X	X			6.7	7.3	11.7	35.5	13.8	28.8	9.2	13.9	23.7	<0.001	0.990	0.180
X		X		7.8	6.3	18.4	32.2	23.7	26.3	9.2	13.1	23.4	<0.001	0.209	0.358
X			X	7.6	7.3	20.1	27.2	23.7	27.6	8.9	12.2	22.7	<0.001	0.138	0.113
	X	X		6.6	5.2	9.8	37.0	17.9	28.6	9.3	14.0	23.7	<0.001	0.674	0.229
	X		X	6.6	3.8	8.9	31.5	9.06	30.7	9.0	12.2	22.9	<0.001	0.755	0.068
		X	X	7.3	3.8	16.7	28.5	18.1	27.0	9.0	11.6	22.5	<0.001	0.516	0.169
X	X	X		6.6	5.2	9.4	41.4	10.2	29.9	9.1	12.7	23.7	<0.001	0.777	0.139
X	X		X	6.6	3.5	8.4	35.6	8.2	30.9	8.9	11.6	23.2	<0.001	0.741	0.082
X		X	X	7.1	3.5	13.6	32.4	16.5	27.1	9.0	11.0	22.8	<0.001	0.591	0.231
	X	X	X	6.4	3.1	7.5	37.1	6.0	30.3	9.0	11.0	23.0	<0.001	0.636	0.115

1587 *Overall observed seroprevalence in 2013 field study was 11.1% in urban, 15.3% in peri-urban, and 24.0% in rural areas.

1588 #Using test of difference between proportions, statistically significant results ($p < 0.05$) in bold.
 1589
 1590
 1591
 1592
 1593
 1594
 1595

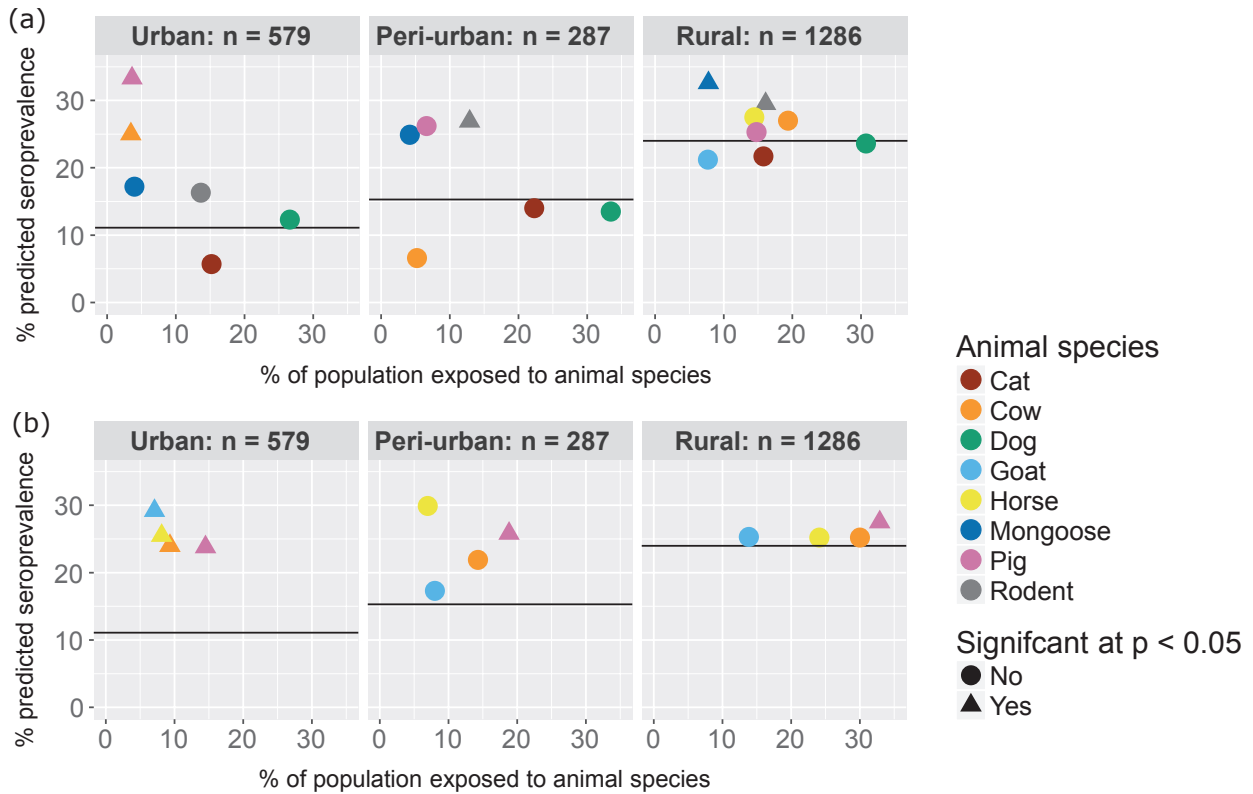


Figure 7. a) Individual/household-level exposure to animals – differences in exposure and predicted seroprevalence between residential settings. Exposure is defined as the presence of animal species at the individual’s household. b) Community-level exposure to animals – differences in exposure and predicted seroprevalence between residential settings. Exposure is defined as the presence of animal species at the individual’s community. Horizontal black lines indicate mean seroprevalence for each subgroup. Triangles/circles indicate a statistically significant/insignificant difference in seroprevalence between exposed and un-exposed groups.

DISCUSSION

Our study showed that model performance of BNs can be significantly improved by building models that better represent causality and account for dependencies among predictor and indicator variables. Group C BNs were structured to better represent causality and complex interdependencies between nodes, and performed better than the naïve BNs in Group A that were conceptually similar to standard regression models (i.e. predictor/indicator variables were independent). Our study demonstrated some useful features of BNs, including the ability to refine model structure, include strongly correlated predictor and indicator variables, and predict outcomes under complex scenarios. We used leptospirosis as a case study, but the approaches presented here could potentially be used to model other diseases or health outcomes.

We determined AUC for our models and obtained values between 0.58 and 0.63 for 25th and 75th percentile limits (Table 8), indicating poor model discrimination power. This means the

1655
1656
1657
1658 probability distributions for true positives and true negatives overlapped significantly, which can
1659 occur in situations of low prevalence, where the model never predicts a high probability for true
1660 cases. This could explain the poor AUC results for our models.³⁷ Also, AUC does not take into
1661 account the predicted probability values and model goodness-of-fit.³⁷ There are many other metrics
1662 that are commonly used to assess the performance of presence/absence models such as the ones
1663 presented in this paper. These include sensitivity (true positive rate), specificity (true negative rate),
1664 the True Skill Statistic (TSS) and Kappa Statistic. We chose to use AUC in our study because it
1665 measures performance across multiple cutoffs, while TSS and Kappa Statistic use the most probable
1666 outcome as the cutoff. The limitation of using the most probable outcome as the cutoff is that the
1667 metric becomes less reliable when prevalence rates (true positive rates) for the outcome being
1668 predicted are low, because in these situations a model trained on data containing low prevalence
1669 rates will rarely predict a high probability of presence for any scenario. In our study, the predicted
1670 and actual seroprevalence of *Leptospira* for any scenario were mostly below 30%, indicating low
1671 prevalence rates.
1672
1673
1674
1675
1676
1677
1678
1679

1680
1681 We compared predicted versus observed seroprevalence to assess the predictive performance of our
1682 BNs and found that Group C models (with the most complex structures) had the best performance,
1683 with R² values of 0.75 to 1.00, and lower MSE compared to Group A and B models. Using the
1684 Group C models, we found that scenario analyses provided important insights into the relative
1685 importance of animal species in leptospirosis transmission in different ethnic groups and residential
1686 settings. These insights were gained by predicting outcomes under complex scenarios that included
1687 multiple correlated predictor/indicator variables, which would have been more difficult to achieve
1688 with regression methods. A central challenge in leptospirosis control is to identify specific points in
1689 exposure pathways where public health and environmental health interventions are likely to be most
1690 effective. Because of the complex and variable transmission dynamics of leptospirosis, scenario
1691 analyses using BNs could be useful for providing insights to inform more targeted prevention and
1692 control strategies for subpopulations.
1693
1694
1695
1696
1697
1698
1699

1700
1701 In the Pacific Islands, leptospirosis has been identified in many animal species including rodents
1702 and livestock,³⁸ and is considered as one of the most important livestock diseases in the region in
1703 terms of impact on human health.³⁹ However, the relative importance of each animal disease in
1704 human infections is currently poorly understood. Our results provide epidemiological evidence that
1705 multiple animal species are likely to be important in leptospirosis transmission in Fiji, and that the
1706 intensity of exposure to animals as well as the relative importance of each animal species vary
1707
1708
1709
1710
1711
1712
1713

1714
1715
1716 significantly between ethnic groups and residential settings, i.e. the prevalence of risk factors vary
1717 significantly between subgroups. For iTaukei, contact with rodents and mongoose and community-
1718 level exposure to livestock were strongly associated with infection. In contrast, very few Indo-
1719 Fijians reported contact with rodents or mongoose, but household exposure to livestock was
1720 important. In rural areas, physical contact with rodents and mongoose were important. Community-
1721 level exposure to pigs was important in all residential settings. Importantly, in urban settings,
1722 exposure to livestock was associated with a very high risk of infection, possibly because animals are
1723 kept closer to homes compared to rural areas. This finding is concerning, because the combination
1724 of population growth, urbanisation, and agricultural intensification (including subsistence farming)
1725 might fuel future urban outbreaks in this setting.
1726
1727

1732 Detailed insights about the role of different animal species in different socio-ecological niches
1733 could potentially be useful for designing interventions that are specifically relevant for subgroups,
1734 e.g. health promotion messages related to contact with rodents and mongoose should be particularly
1735 strengthened in iTaukei communities, but improving management of livestock animals is important
1736 for all communities in Fiji. Animal and anthropological studies will be required to confirm the
1737 epidemiological associations identified by our study. Our findings provide important baseline data
1738 for developing future studies to assess the impact of interventions in Fiji, e.g. evaluating specific
1739 strategies for each ethnic group and residential setting.
1740
1741

1742 Our results should be interpreted in light of the study's limitations. The study's outcome measure
1743 was the presence of *Leptospira* antibodies, which is an indication of prior infection. However, many
1744 leptospirosis infections are asymptomatic and the severity of clinical infections depends on many
1745 factors including age, comorbidities, and pathogen virulence. Our study used animal data at the
1746 place of residence, but it is possible for infections to occur elsewhere. The database was obtained
1747 from a cross-sectional study conducted in 2013, and it is possible for risk factors to evolve over
1748 time.
1749

1750 The application of BNs in infectious disease epidemiology has recently been increasing. A recent
1751 study used BNs to model meningitis outbreaks in the Niger using historical epidemiological
1752 databases, and concluded that BNs provide a promising approach for understanding the dynamics of
1753 epidemics, estimating the risk of outbreaks, and providing information to target control
1754 interventions.⁴⁰ BNs have also been used to model seasonal and population influences on
1755 pneumonic plague,⁴¹ the impact of demographics and vaccination on influenza,⁴² hierarchical
1756
1757
1758
1759
1760
1761
1762
1763
1764
1765
1766
1767
1768
1769
1770
1771
1772

1773
1774
1775
1776 relationships of risk factors associated with infectious diarrhoea in children,⁴³ and household factors
1777 that influence the risk of malaria in sub-Saharan Africa.⁴⁴
1778

1779
1780 There are other advantages of Bayesian networks that were not fully explored in this study,
1781 including their graphical user interface that allow models to be more easily understood and
1782 interpreted; the interactive and dynamic setup that allows users to define complex scenarios and see
1783 updated predictions almost immediately; the ability to incorporate different sources and types of
1784 knowledge including empirical data and expert opinion; the ease with which new data can be
1785 incorporated into models to update probabilistic relationships between variables; the ability to
1786 model causal pathways; and the ability to use the models in predictive or diagnostic modes, or a
1787 combination of both as shown in the BNs used in this study.^{11,12,33} BNs have therefore been used in
1788 many disciplines including medicine, ecology, environmental sciences, engineering, gaming, and
1789 artificial intelligence.
1790
1791
1792
1793
1794
1795

1796
1797
1798 Future work on the use of BNs in infectious disease epidemiology should explore the development
1799 of more complex models that incorporate a wider range or predictor/indicator variables, including
1800 variables that operate at different ecological scales. Integrating BNs with other types of models that
1801 include spatial, temporal or dynamic components will also help improve understanding of disease
1802 transmission.
1803
1804
1805
1806
1807

1808 **CONCLUSIONS**

1809

1810
1811 We demonstrated that BNs provide a promising alternative approach to modelling infectious disease
1812 epidemiology and unravelling the complex drivers of transmission. Using BNs, our study provided
1813 important information on the role of different animal species in leptospirosis transmission in Fiji.
1814 We showed that the drivers of leptospirosis transmission are likely to vary between socio-ecological
1815 niches, with important implications for targeted prevention and control strategies. Although our
1816 study focused on leptospirosis in Fiji, the analytical approaches could be used to model other
1817 diseases or health outcomes.
1818
1819
1820
1821
1822
1823
1824
1825
1826
1827
1828
1829
1830
1831

1832
1833
1834
1835 **AUTHOR CONTRIBUTIONS**

1836 CL and CS conceived the study and proposed the use of Bayesian networks; CL, HM, and CS
1837 designed the study, analysed the data, and developed the figures and tables. CW, CL, MK, and EN
1838 contributed to the design of the field study and collection of survey data. JL, CL, and CW
1839 integrated spatial environmental and socio-demographic data with the survey data. CL, HM, and CS
1840 prepared the draft manuscript. All authors reviewed and approved the final manuscript.
1841
1842
1843
1844

1845 **ACKNOWLEDGEMENTS**

1846 We extend our warmest thanks to the many participants, communities, Ministry of Health staff and
1847 community health workers who generously contributed to the field study conducted in Fiji in 2013.
1848 We also thank the WHO/FAO/OIE Collaborating Centre for Reference and Research on
1849 Leptospirosis (at Health Support Queensland, Department of Health, Brisbane, Australia) for
1850 providing the laboratory support for the study.
1851
1852
1853
1854

1855 **CONFLICTS OF INTEREST:**

1856 The authors do not have any conflicts of interest to declare.
1857
1858

1859 **FUNDING SOURCES:**

1860 CLL was supported by an Australian National Health and Medical Research Council (NHMRC)
1861 Fellowship (1109035). CHW was supported by the UK Medical Research Council (grant
1862 MR/J003999/1) and the Chadwick Trust. The study was partly funded by a University of
1863 Queensland Early Career Researcher Grant (2014003059). The funders had no role in study design,
1864 data collection and analysis, decision to publish, or preparation of the manuscript.
1865
1866
1867
1868
1869
1870
1871
1872
1873
1874
1875
1876
1877
1878
1879
1880
1881
1882
1883
1884
1885
1886
1887
1888
1889
1890

1891
1892
1893
1894
1895
1896
1897
1898
1899
1900
1901
1902
1903
1904
1905
1906
1907
1908
1909
1910
1911
1912
1913
1914
1915
1916
1917
1918
1919
1920
1921
1922
1923
1924
1925
1926
1927
1928
1929
1930
1931
1932
1933
1934
1935
1936
1937
1938
1939
1940
1941
1942
1943
1944
1945
1946
1947
1948
1949

REFERENCES

1. Eisenberg JN, 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.
2. Rabinowitz PM, et al. Toward Proof of Concept of a One Health Approach to Disease Prediction and Control. *Emerg Infect Dis.* 2013;19:e130265.
3. Mills JN, et al. Potential Influence of Climate Change on Vector-Borne and Zoonotic Diseases: A Review and Proposed Research Plan. *Environ Health Perspect.* 2010;118:1507-14.
4. Jones KE, et al. Global Trends in Emerging Infectious Diseases. *Nature.* 2008;451:990-3.
5. McMichael AJ. Prisoners of the Proximate: Loosening the Constraints on Epidemiology in an Age of Change. *Am J Epidemiol.* 1999;149:887-97.
6. Susser M. Does Risk Factor Epidemiology Put Epidemiology at Risk? Peering into the Future. *J Epidemiol Community Health.* 1998;52:608-11.
7. Burns CJ, et al. Evaluating Uncertainty to Strengthen Epidemiologic Data for Use in Human Health Risk Assessments. *Environ Health Perspect.* 2014;122:1160-5.
8. Lau CL, et al. Bayesian Networks in Infectious Disease Eco-Epidemiology. *Rev Environ Health.* 2016.
9. Keyes KM, et al. Commentary: The Limits of Risk Factors Revisited: Is It Time for a Causal Architecture Approach? *Epidemiology.* 2017;28:1-5.
10. Poole C. Commentary: Some Thoughts on Consequential Epidemiology and Causal Architecture. *Epidemiology.* 2017;28:6-11.
11. Joffe M, et al. Causal Diagrams in Systems Epidemiology. *Emerg Themes Epidemiol.* 2012;9:1.
12. Landuyt D, et al. A Review of Bayesian Belief Networks in Ecosystem Service Modelling. *Environ Model Softw.* 2013;46:1-11.

1950
1951
1952
1953
1954
1955
1956
1957
1958
1959
1960
1961
1962
1963
1964
1965
1966
1967
1968
1969
1970
1971
1972
1973
1974
1975
1976
1977
1978
1979
1980
1981
1982
1983
1984
1985
1986
1987
1988
1989
1990
1991
1992
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008

13. Costa F, et al. Global Morbidity and Mortality of Leptospirosis: A Systematic Review. *PLoS Negl Trop Dis*. 2015;9:e0003898.
14. Bharti AR, et al. Leptospirosis: A Zoonotic Disease of Global Importance. *Lancet Infect Dis*. 2003;3:757-71.
15. Lau CL, et al. Climate Change, Flooding, Urbanisation and Leptospirosis: Fuelling the Fire? *Trans R Soc Trop Med Hyg*. 2010;104:631-8.
16. Mwachui MA, et al. Environmental and Behavioural Determinants of Leptospirosis Transmission: A Systematic Review. *PLoS Negl Trop Dis*. 2015;9:e0003843.
17. Jones BA, et al. Zoonosis Emergence Linked to Agricultural Intensification and Environmental Change. *Proc Natl Acad Sci*. 2013;110:8399-404.
18. Hartskeerl RA, et al. Emergence, Control and Re-Emerging Leptospirosis: Dynamics of Infection in the Changing World. *Clin Microbiol Infect*. 2011;17:494-501.
19. Coker R, et al. Towards a Conceptual Framework to Support One-Health Research for Policy on Emerging Zoonoses. *Lancet Infect Dis*. 2011;11:326-31.
20. Liverani M, et al. Understanding and Managing Zoonotic Risk in the New Livestock Industries. *Environ Health Perspect*. 2013;121:873-7.
21. McIver L, et al. Health Impacts of Climate Change in Pacific Island Countries: A Regional Assessment of Vulnerabilities and Adaptation Priorities. *Environ Health Perspect*. 2015.
22. Averett N. Pacific Island Countries and Climate Change: Examining Associated Human Health Vulnerabilities. *Environ Health Perspect*. 2016;124:A208.
23. Kline K, et al. Neglected Tropical Diseases of Oceania: Review of Their Prevalence, Distribution, and Opportunities for Control. *PLoS Negl Trop Dis*. 2013;7:e1755.
24. Weinberger D, et al. El Nino Southern Oscillation and Leptospirosis Outbreaks in New Caledonia. *PLoS Negl Trop Dis*. 2014;8:e2798.

2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
2023
2024
2025
2026
2027
2028
2029
2030
2031
2032
2033
2034
2035
2036
2037
2038
2039
2040
2041
2042
2043
2044
2045
2046
2047
2048
2049
2050
2051
2052
2053
2054
2055
2056
2057
2058
2059
2060
2061
2062
2063
2064
2065
2066
2067

25. Massenet D, et al. An Unprecedented High Incidence of Leptospirosis in Futuna, South Pacific, 2004 - 2014, Evidenced by Retrospective Analysis of Surveillance Data. *PLoS One*. 2015;10:e0142063.
26. Colt S, et al. Human Leptospirosis in the Federated States of Micronesia: A Hospital-Based Febrile Illness Survey. *BMC Infect Dis*. 2014;14:186.
27. Lau C, et al. Human Leptospirosis Infection in Fiji: An Eco-Epidemiological Approach to Identifying Risk Factors and Environmental Drivers for Transmission. *PLoS Negl Trop Dis*. 2016;10:e0004405.
28. Lau CL, et al. Leptospirosis in American Samoa - Estimating and Mapping Risk Using Environmental Data. *PLoS Neglected Tropical Diseases*. 2012;6:e1669.
29. Ram P, et al. Further Observations on the Epidemiology of Leptospirosis in Fiji. *Fiji Medical Journal*. 1982;10:71-5.
30. Ghosh A, et al. Leptospirosis in Fiji: Incidence between 2000 to 2007 and Review of Literature. *Fiji Medical Journal*. 2010;29:8-14.
31. McIver L, et al. Climate Change and Health in Fiji: Environmental Epidemiology of Infectious Diseases & Potential for Climate-Based Early Warning Systems. *Fiji Journal of Public Health*. 2012;1:7-13.
32. Fiji Bureau of Statistics. Census of Population and Housing. 2007. <http://www.statsfiji.gov.fj/index.php/2007-census-of-population>. Accessed 8 September 2015.
33. Fenton N, et al. Risk Assessment and Decision Analysis with Bayesian Networks. Boca Raton: CRC Press, Taylor & Francis Group.; 2013.
34. Norsys Software Corp. Netica Bayesian Belief Network Software. <https://http://www.norsys.com2013>.
35. Fenton N, et al. Chapter 7. Defining the Structure of Bayesian Networks. Risk Assessment and Decision Analysis with Bayesian Networks. Boca Raton: CRC Press, Taylor & Francis Group.; 2013.

- 2068
2069
2070
2071
2072
2073
2074
2075
2076
2077
2078
2079
2080
2081
2082
2083
2084
2085
2086
2087
2088
2089
2090
2091
2092
2093
2094
2095
2096
2097
2098
2099
2100
2101
2102
2103
2104
2105
2106
2107
2108
2109
2110
2111
2112
2113
2114
2115
2116
2117
2118
2119
2120
2121
2122
2123
2124
2125
2126
36. Do CB, et al. What Is the Expectation Maximization Algorithm? *Nat Biotechnol.* 2008;26:897-9.
 37. Lobo JM, et al. Auc: A Misleading Measure of the Performance of Predictive Distribution Models. *Global Ecology and Biogeography.* 2007;17:145-51.
 38. Brioudes A, et al. A Review of Domestic Animal Diseases within the Pacific Islands Region. *Acta Trop.* 2014;132:23-38.
 39. Brioudes A, et al. Diseases of Livestock in the Pacific Islands Region: Setting Priorities for Food Animal Biosecurity. *Acta Trop.* 2015;143:66-76.
 40. Beresniak A, et al. A Bayesian Network Approach to the Study of Historical Epidemiological Databases: Modelling Meningitis Outbreaks in the Niger. *Bull World Health Organ.* 2012;90:412-7A.
 41. Dawson P, et al. Detecting Disease Outbreaks Using a Combined Bayesian Network and Particle Filter Approach. *Journal of Theoretical Biology.* 2015;370:171-83.
 42. Abbas K, et al. Temporal Analysis of Infectious Diseases: Influenza. Proceedings of the 2005 ACM symposium on Applied computing. Santa Fe, New Mexico: ACM; 2005:267-71.
 43. Nguefack-Tsague G. Using Bayesian Networks to Model Hierarchical Relationships in Epidemiological Studies. *Epidemiology and Health.* 2011;33:e2011006.
 44. Semakula HM, et al. A Bayesian Belief Network Modelling of Household Factors Influencing the Risk of Malaria: A Study of Parasitaemia in Children under Five Years of Age in Sub-Saharan Africa. *Environmental Modelling & Software.* 2016;75:59-67.