Acta Tropica 173 (2017) 191–199

Contents lists available at ScienceDirect

Acta Tropica

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

Comparison between generalized linear modelling and additive Bayesian network; identification of factors associated with the incidence of antibodies against *Leptospira interrogans* sv *Pomona* in meat workers in New Zealand

M. Pittavino[a,1,2,*], A. Dreyfus[b,1], C. Heuer[c], J. Benschop[c], P. Wilson[c], J. Collins-Emerson[c], P.R. Torgerson[b], R. Furrer[a,d]

[a] Department of Mathematics, University of Zurich, Zurich, Switzerland
[b] Section of Epidemiology, Vetsuisse Faculty, University of Zurich, Zurich, Switzerland
[c] Institute of Veterinary Animal and Biomedical Sciences, Massey University, Palmerston North, New Zealand
[d] Department of Computational Science, University of Zurich, Zurich, Switzerland

ARTICLE INFO

Keywords:
Leptospirosis
MCMC
R
JAGS
Risk factors
Protective equipment

ABSTRACT

Background: Additive Bayesian Network (ABN) is a graphical model which extends Generalized Linear Modelling (GLM) to multiple dependent variables. The present study compares results from GLM with those from ABN analysis used to identify factors associated with *Leptospira interrogans* sv Pomona (Pomona) infection by exploring the advantages and disadvantages of these two methodologies, to corroborate inferences informing health and safety measures at abattoirs in New Zealand (NZ).

Methodology and findings: In a cohort study in four sheep slaughtering abattoirs in NZ, sera were collected twice a year from 384 meat workers and tested by Microscopic Agglutination with a 91% sensitivity and 94% specificity for Pomona.

The study primarily addressed the effect of work position, personal protective equipment (PPE) and non-work related exposures such as hunting on a new infection with Pomona. Significantly associated with Pomona were "Work position" and two "Abattoirs" (GLM), and "Work position" (ABN). The odds of Pomona infection (OR, [95% CI]) was highest at stunning and hide removal (ABN 41.0, [6.9–1044.2]; GLM 57.0, [6.9–473.3]), followed by removal of intestines, bladder, and kidneys (ABN 30.7, [4.9–788.4]; GLM 33.8, [4.2–271.1]). Wearing a facemask, glasses or gloves (PPE) did not result as a protective factor in GLM or ABN.

Conclusions/Significance: The odds of Pomona infection was highest at stunning and hide removal. PPE did not show any indication of being protective in GLM or ABN. In ABN all relationships between variables are modelled; hence it has an advantage over GLM due to its capacity to capture the natural complexity of data more effectively.

1. Introduction

The present study compares results from Generalized Linear Modelling (GLM) with those from Additive Bayesian Network (ABN) analysis by exploring the advantages and disadvantages of these two analytical methods while analysing risk factors for occupational leptospirosis in New Zealand (NZ).

A primary objective of many epidemiological studies is to investigate hypothesized relationships between covariates of interest, and one or more outcome variables. To date, a large variety of statistical models is available to analyse epidemiological data (i.e. cross validation criteria, ANOVA), and one of the most popular is GLM (McCulloch et al., 2008). Typically, the biological and epidemiological processes, which generated these data, are highly complex, resulting in multiple correlations/dependencies between covariates and also between outcome variables. Standard epidemiological and statistical approaches have a limited ability to describe such inter-dependent multi-factorial relationships. ABN is a form of probabilistic graphical model that extends the usual GLM to multiple dependent variables, through the representation of the joint probability distribution of random variables. It is a statistical model that allows the analysis of complex data and derives a directed acyclic graph (DAG) from empirical data, describing the

* Corresponding author. Department of Mathematics, University of Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland.
E-mail address: marta.pittavino@math.uzh.ch (M. Pittavino).
[1] Joint first authors.
[2] Now at the International Agency for Research on Cancer, Lyon, France.

http://dx.doi.org/10.1016/j.actatropica.2017.04.034
Received 4 November 2016; Received in revised form 13 April 2017; Accepted 14 April 2017
Available online 06 May 2017

dependency structure between random variables as opposed to fixed variables in GLM (Sivasundaram, 2012; Rijmen, 2008). ABN models comprise two reciprocally dependent parts: a DAG and a set of parameters. A DAG is a graphical representation of the joint probability distribution of all random variables in the data. Each node in the DAG is the equivalent to the dependent variable in a GLM regression model. In a graphical statistical model there is no distinction between covariates and an outcome variable. Hence, while a standard GLM focuses on the association between covariates and a single dependent or outcome variable, an ABN is a multivariate (conditional) regression model, analysing the associations between all covariates with all variables being potentially dependent (Lewis and McCormick, 2012). Therefore, in a multifactorial complex disease system, interdependencies between risk factors may be revealed in ABN, that may or may not be discovered in GLM, as the latter imposes a linear relationship between covariates and the outcome (Lewis and McCormick, 2012). By comparing ABN with GLM using identical data, we explore the likely impact of such an analytical difference on the inferences from this study.

The ABN models described here, also if consisting of a DAG, are only related with statistical dependency, and arcs present in such models do not imply any causal relationship. While the identification of a statistical dependency is often a step towards the conclusion of causal mechanisms, it is, however, more demanding to further claim that the given dependency exists within a particular causal web.

In the last decades, Bayesian Network (BN) modelling has been widely used in biomedical science/systems biology (Lycett et al., 2009; Poon et al., 2008; Poon et al., 2007a; Poon et al., 2007b; Dojer et al., 2006; Hodges et al., 2010; Jansen et al., 2003; Needham et al., 2007; Djebbari and Quackenbush, 2008) to analyse multi-dimensional data. However, only in the last few years, it has been applied in the veterinary epidemiology field. A general introduction to BN modelling in veterinary epidemiology is provided by Lewis et al. (2011). Further applications of BN to veterinary studies were described by Ward and Lewis (2013), Wilson et al. (2013), Sanchez-Vazquez et al. (2012). Graphical modelling techniques used to analyse epidemiological data were used by Firestone et al. (2013), Schemann et al. (2013), Lewis et al. (2011), Ludwig et al. (2013) and McCormick et al. (2013). Some of these do not compare results from ABN and GLM (Firestone et al., 2013; Firestone et al., 2014; Schemann et al., 2013), whereas others do (Sivasundaram, 2012; Ludwig et al., 2013; McCormick et al., 2013; Lewis and Ward, 2013). In the literature, a detailed comparison of these two methodologies can be found in Lewis and Ward (2013). However, the aforementioned study was based on simulated (artificial) epidemiological data and differences between results were mainly discussed with graphical outputs (qualitatively), whereas this analysis also compares ORs of parameters directly and indirectly linked to the outcome, focusing on the contrast as well on a quantitative point of view. “Additive” BN models have the advantage over the “classical” BN in allowing a direct comparison between the reciprocal model parameters. While BN parameters are based on contingency tables, the resulting data counts ABN refers to regression parameters resulting from the transformation through a link function (here logit) of the cell probability parameters. Hence, ABNs are more appropriate and suitable for the aim of the presented work.

Leptospirosis is a zoonotic disease occurring in many mammals and is caused by a bacterium of the genus *Leptospira* spp. Transmission occurs from exposure to urine or aborted tissues of infected animals, either directly or via contact with contaminated water or soil (Hartskeerl et al., 2011). Pathogenic leptospires enter the body through mucous membranes or skin abrasions. In humans, infection with *Leptospira* spp. varies from being sub-clinical (asymptomatic), through a mild to a severe acute disease. A mild form with fever and “influenza-like” symptoms appears to be more common in New Zealand (Dreyfus et al., 2014a). The acute disease is characterized by jaundice, renal failure, hepatic failure, myocarditis, uveitis and/or pulmonary haemorrhage (Adler, 2010; Bharti et al., 2003).

Among temperate developed countries, New Zealand (NZ) has a relatively high incidence of notified human leptospirosis cases with an average annual incidence risk of 2–3 cases per 100,000 population (Thornley et al., 2002; ESR, 2010). However, under-ascertainment is common and estimated to be 15–65 fold in sheep abattoir workers (Dreyfus et al., 2014a). The three most common serovars in humans are *Leptospira interrogans* sv Pomona (Pomona) and *Leptospira borgpetersenii* sv Hardjo (Hardjo) and *Leptospira interrogans* sv Ballum (Ballum) (ESR, 2010). The serovar Pomona is highly prevalent in cattle, deer and sheep in NZ (Dreyfus, 2013; Marshall and Manktelow, 2002; Ayanegui-Alcerreca et al., 2010). Therefore, livestock are a frequent source of human leptospirosis in farmers and meat workers (Thornley et al., 2002) who are most at risk with less than 10% of deer mobs, sheep flocks or beef herds currently vaccinated against leptospirosis (Wilson et al., 2008; Keenan, 2007). Dreyfus et al. (Dreyfus et al., 2014a) found that in 2011 the annual cumulative Pomona incidence risk (%) in sheep abattoir workers was on average 11.9% with a range for four different abattoirs of 8.4-16.4%. The annual risk of confirmed clinical leptospirosis was 0.78% (3/384, 95% CI 0.20-2.46%) and new infections with Pomona increased the risk of illness with ‘influenza-like’ symptoms 2.1-fold (Dreyfus et al., 2014a).

This study used the data of the study described above (Dreyfus et al., 2014a) with the following two aims: the first aim was to identify factors associated with Pomona infection in sheep abattoir workers in NZ, with two different methodologies GLM and ABN, in order to untangle the web of causality of human infection with Pomona with a real data set. Specifically, we aimed to test the hypothesis of work position being a strongly associated variable, to evaluate the role of personal protective equipment (PPE) and non-work related exposures, such as hunting, home slaughtering and farming. If PPE had a protective effect, it would be a good measure to protect workers. If workers were mainly exposed in their work place and not while hunting or home slaughtering, then it becomes clear where the emphasis on their protection should be. The second and equally important aim was to compare the results between GLM and ABN and discuss advantages and disadvantages of the two statistical analyses.

2. Materials and methods

2.1. Case study

A prospective cohort study amongst voluntarily participating meat workers from four purposively selected sheep abattoirs in the North Island of NZ was conducted. Study methods were described in detail by Dreyfus et al. (2014a). Participants were blood sampled by certified phlebotomists or nurses and interviewed at the same time by trained researchers using a questionnaire (Supplementary Material). Serum antibodies against Pomona were analysed by the microscopic agglutination test (MAT) at doubling dilutions from 1:24 to 1:1536 as described previously (Faine et al., 1999). Blood samples and data were collected twice at intervals ranging from 50 to 61 weeks in order to estimate the incidence of new infections with Pomona. Study participants of “Abattoir 1” were sampled the first time between February and April 2008 and the second time in April 2009. All other abattoirs were sampled initially in November 2009 − March 2010, and again in November 2010 − May 2011. Hence, one abattoir (“Abattoir 1”) was studied twice in two consecutive years and three abattoirs were studied in the second year once. New infection occurred where a worker seroconverted (a sero-negative worker had a MAT titre increase to equal or higher than 1:48) or had an anamnestic response (a sero-positive worker had a MAT titre increase by two or more dilutions) (Dreyfus et al., 2014a).

2.2. Data structure

Serological test results and questionnaire information were entered