



Zero-inflated models for identifying disease risk factors when case detection is imperfect: Application to highly pathogenic avian influenza H5N1 in Thailand



Timothée Vergne^{a,b,c,*}, Mathilde C. Paul^{a,d}, Wanida Chaengprachak^e, Benoit Durand^b, Marius Gilbert^{f,g}, Barbara Dufour^h, François Roger^a, Suwicha Kasemsuwanⁱ, Vladimir Grosbois^a

^a AGIRs Unit (UR22), Centre de Coopération Internationale en Recherche Agronomique pour le Développement, Montpellier, France

^b Laboratoire de Santé Animale, Agence de Sécurité Sanitaire, Maisons-Alfort, France

^c Veterinary Epidemiology Economics and Public Health, Royal Veterinary College, London, United Kingdom

^d Université de Toulouse, INP-ENVT, INRA UMR 1225 IHAP, Toulouse, France

^e Department of Livestock Development, Bangkok, Thailand

^f Université Libre de Bruxelles, Bruxelles, Belgium

^g Fonds National de la Recherche Scientifique, Bruxelles, Belgium

^h Ecole Nationale Vétérinaire d'Alfort, Maisons-Alfort, France

ⁱ Faculty of Veterinary Medicine, Kasetsart University, Bangkok, Thailand

ARTICLE INFO

Keywords:

Zero-inflation
Surveillance
Under-detection
Capture–recapture
Risk factors
Evaluation
Avian influenza H5N1
Bias
Count
Spatial
Conditional autoregressive model

ABSTRACT

Logistic regression models integrating disease presence/absence data are widely used to identify risk factors for a given disease. However, when data arise from imperfect surveillance systems, the interpretation of results is confusing since explanatory variables can be related either to the occurrence of the disease or to the efficiency of the surveillance system. As an alternative, we present spatial and non-spatial zero-inflated Poisson (ZIP) regressions for modelling the number of highly pathogenic avian influenza (HPAI) H5N1 outbreaks that were reported at subdistrict level in Thailand during the second epidemic wave (July 3rd 2004 to May 5th 2005). The spatial ZIP model fitted the data more effectively than its non-spatial version. This model clarified the role of the different variables: for example, results suggested that human population density was not associated with the disease occurrence but was rather associated with the number of reported outbreaks given disease occurrence. In addition, these models allowed estimating that 902 (95% CI 881–922) subdistricts suffered at least one HPAI H5N1 outbreak in Thailand although only 779 were reported to veterinary authorities, leading to a general surveillance sensitivity of 86.4% (95% CI 84.5–88.4). Finally, the outputs of the spatial ZIP model revealed the spatial distribution of the probability that a subdistrict could have been a false negative. The methodology presented here can easily be adapted to other animal health contexts.

© 2014 Elsevier B.V. All rights reserved.

1. Introduction

The first outbreaks of highly pathogenic avian influenza (HPAI) H5N1 were reported in Thailand in January 2004. For around two years, a large epidemic occurred throughout the country. Most outbreaks were located in the Central

* Corresponding author at: AGIRs Unit (UR22), Centre de Coopération Internationale en Recherche Agronomique pour le Développement, Montpellier, France.

E-mail address: timotheevergne@yahoo.com (T. Vergne).

Plain, causing massive mortality in chickens and ducks (Tiensin et al., 2005). The economic consequences of these outbreaks were dramatic, as more than 65 million birds were culled and over US\$ 130 million was spent compensating farmers' losses during 2004 and 2005 (Tiensin et al., 2007). However, the number of outbreaks significantly decreased late in the first half of 2005, and since then, HPAI H5N1 has only caused sporadic outbreaks in the country.

It is essential to study the risk factors of highly pathogenic avian influenza (HPAI) H5N1 in order to design suitable surveillance and control strategies. These risk factors have been extensively studied over the past few years, mainly in Southeast Asia but also in Africa and Europe (Gilbert and Pfeiffer, 2012). Most of these studies have defined their epidemiologic unit as an administrative unit, such as the subdistrict in Thailand (Gilbert et al., 2006; Tiensin et al., 2009; Paul et al., 2010) or the commune in Vietnam (Pfeiffer et al., 2007; Henning et al., 2009; Minh et al., 2009) and have modelled the disease's presence/absence through logistic regression models. However, Gilbert and Pfeiffer (2012) pointed out that variables significantly associated with the presence of HPAI H5N1 in logistic regressions could appear significant simply because they have a positive influence on the sensitivity of detection rather than on the processes driving disease presence/absence. It therefore appears essential to account for imperfect case detection when identifying disease risk factors.

The alternative methodology used in this paper follows the general approach introduced in site occupancy studies in ecology (MacKenzie et al., 2002, 2006). Site occupancy models estimate the proportion of sites occupied by a species, although the probability of detecting the species at each detection occasion on each site is less than one. In disease surveillance where there are no formal detection occasions, data can no longer be viewed as a series of detections/non-detections, and should instead be considered in terms of the number of detections on each site. Zero-inflated models are commonly used in such applications (Martin et al., 2005; Royle et al., 2005).

Zero-inflated regressions were introduced by Lambert (1992) and subsequently applied in various contexts such as social science (Heibron, 1994), spatial analysis (Agarwal et al., 2002), ecology (Martin et al., 2005; Flores et al., 2009), public health (Böhning et al., 1997; Musal and Aktekin, 2013), and veterinary epidemiology (Lockhart, 2008; Benschop et al., 2010; Bronner et al., 2013). Zero inflation occurs when count data display an excessive frequency of zeros, inconsistent with usual count distributions such as Poisson or negative binomial distributions (Cameron and Trivedi, 1998). In ecology, empirical investigations of site occupancy patterns of a focal species often reveal excess zeros in distributions of counts of detection of the focal species over study sites. One explanation of this phenomenon is that sites where the species is present but never detected generate additional "false zeros" (Martin et al., 2005).

In Thailand, the distribution of the number of HPAI H5N1 outbreaks reported in each subdistrict during the second epidemic wave is likely to be structurally

zero-inflated. A subdistrict where no outbreaks were detected could indeed be a subdistrict where no outbreaks occurred (true zero), but it could also be a subdistrict where at least one outbreak occurred but none were reported (false zero). Therefore, in order to correctly model HPAI H5N1 surveillance count data from Thailand at subdistrict level, we used the zero-inflated approach suggested by Lockhart (2008, chapter 5). Note that zero-inflated models allow for the presence of false negatives (sensitivity of the surveillance < 1), but assume the absence of false positives (specificity of the surveillance = 1). In the online supplementary material 1, we present the results of a simulation study demonstrating the limitations of logistic models to identify risk factors of disease presence when case detection is imperfect. These simulations also illustrate the potential of zero inflated models for drawing inference on risk factors associated with the occurrence of a disease when count data blurred by imperfect detection are available.

In this paper, we used a spatial version of the zero-inflated Poisson model to model the number of HPAI H5N1 outbreaks reported at subdistrict level in Thailand during the second epidemic wave from July 3rd 2004 to May 5th 2005. The objectives of our study were (i) to simultaneously identify the factors associated with disease presence and those associated with the number of reported outbreaks given disease presence, (ii) to estimate the overall sensitivity of HPAI H5N1 surveillance at subdistrict level in Thailand, (iii) to describe the spatial distributions at subdistrict level of the risk of HPAI H5N1 presence and of detection sensitivity, and (iv) to describe the spatial distribution of the probability that a subdistrict was a false zero during the second HPAI H5N1 epidemic wave in Thailand.

2. Materials and methods

2.1. Spatial zero-inflated model

Because zero-inflated regression models are used in epidemiological studies for modelling multiple detections of infected units with the objective of estimating the number of undetected infected units, they can be viewed as a type of capture–recapture approach (Bronner et al., 2013). More specifically, since the data are derived from only one detection source, they fall into the unilist capture–recapture category, just like zero-truncated regression models (Del Rio Vilas and Bohning, 2008; Vergne et al., 2012; Bronner et al., 2013).

In a zero-inflated Poisson model (ZIP), the distribution of observed count Y follows a mixture of a zero-point mass distribution and a Poisson distribution (Cameron and Trivedi, 1998): the model assigns a mass Φ to the Poisson distribution and a mass $(1 - \Phi)$ to the extra zeros. The ZIP (Φ, λ) probability function can be expressed thus:

$$P(Y_i = y | \Phi_i, \lambda_i) = \begin{cases} (1 - \Phi_i) + \Phi_i e^{-\lambda_i} & \text{if } y = 0 \\ \Phi_i \frac{\lambda_i^y e^{-\lambda_i}}{y!} & \text{if } y > 0 \end{cases}$$

where λ is the parameter of the Poisson distribution, $\text{logit}(\Phi_i) = \alpha_0 + \sum_i a_i x_{1i}$ and $\log(\lambda_i) = \beta_0 + \sum_i \beta_0 + x_{2i}$ with

Download English Version:

<https://daneshyari.com/en/article/2452495>

Download Persian Version:

<https://daneshyari.com/article/2452495>

[Daneshyari.com](https://daneshyari.com)