Preventive Veterinary Medicine 110 (2013) 190–197

Contents lists available at SciVerse ScienceDirect

Preventive Veterinary Medicine

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

# Time series analysis of human and bovine brucellosis in South Korea from 2005 to 2010

Hu Suk Lee[a], Moon Her[b], Michael Levine[c], George E. Moore[a,*]

[a] Department of Comparative Pathobiology, College of Veterinary Medicine, Purdue University, 725 Harrison Street, West Lafayette, IN 47907, United States
[b] OIE Reference Laboratory for Brucellosis, Bacterial Disease Division, Department of Animal and Plant Health Research, Animal, Plant and Fisheries Quarantine and Inspection Agency (QIA), 175 Anyang-ro, Manan-gu, Anyang-si, Gyeonggi 430-757, South Korea
[c] Department of Statistics, Purdue University, 250 N. University Street, West Lafayette, IN 47907, United States

ARTICLE INFO

*Article history:*
Received 18 July 2012
Received in revised form
12 November 2012
Accepted 1 December 2012

*Keywords:*
Brucellosis
Time series
Lag
ARIMAX model
NBR model

ABSTRACT

Brucellosis is considered to be one of the most important zoonotic diseases in the world, affecting underdeveloped and developing countries. The primary purpose of brucellosis control is to prevent the spread of disease from animals (typically ruminants) to humans. The main objective of this study was to retrospectively develop an appropriate time series model for cattle-to-human transmission in South Korea using data from independent national surveillance systems. Monthly case counts for cattle and people as well as national population data were available for 2005–2010. The temporal relationship was evaluated using an autoregressive integrated moving average with exogenous input (ARIMAX) model [notated as ARIMA($p$, $d$, $q$) – AR($p$)] and a negative binomial regression (NBR) model.

Human incidence rate was highly correlated to cattle incidence rate in the same month and the previous month (both $r$ = 0.82). In the final models, ARIMA (0, 1, 1) – AR (0, 1) was determined as the best fit with 191.5% error in the validation phase, whereas the best NBR model including lags (0, 1 months) for the cattle incidence rate yielded a 131.9% error in the validation phase. Error (MAPE) rates were high due to small absolute human case numbers (typically less than 10 per month in the validation phase). The NBR model however was able to demonstrate a marked reduction in human case immediately following a hypothetical marked reduction in cattle cases, and may be better for public health decision making.



## 1. Introduction

Brucellosis is considered to be one of the most important zoonotic diseases by the World Health Organization (WHO), Food and Agriculture Organization (FAO) and the World Organization for Animal Health (OIE) (Joint FAO/WHO, 1986; Schelling et al., 2003). Infection with *Brucella abortus* in cattle causes abortions, infertility, and reduced milk production and can cause septicemic and/or granulomatous disease in humans (Halling and Boyle, 2002; Seleem et al., 2010). The primary objective of brucellosis control is to prevent human infections via disease control or eradication in animals. Humans can be easily infected with the *Brucella* organism through direct contact with milk, blood, tissue, or body fluids related to abortion in infected animals. The consumption of unpasteurized milk and cheese has historically been a major source of human infection in many countries (Olsen and Tatum, 2010). The onset of clinical signs in humans is generally a week or month after contact with infected animals or materials, although some infections cause minimal clinical illness (Young, 1983; WHO, 2006). Occupations with animal contact, such as farm workers, veterinarians, ranchers, abattoir workers and lab workers are classified as high risk groups (Seleem et al., 2010). Direct human-to-human transmission rarely occurs, although it has been reported

* Corresponding author. Tel.: +1 765 496 3393; fax: +1 765 494 9830.
*E-mail address:* gemoore@purdue.edu (G.E. Moore).

0167-5877/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.prevetmed.2012.12.003

that transmission may occur via breast-feeding and sexual contact (Arroyo Carrera et al., 2006; Kato et al., 2007). Disease control in humans is therefore accomplished by disease control in animals.

The South Korean government in their Infectious Disease Prevention and Control Act designated brucellosis as a reportable disease in both humans and animals (Kakoma et al., 2007; Wee et al., 2008). The first case of bovine brucellosis in the country was reported among imported dairy cattle in 1955 (Park and Lee, 1959). There has been a steady increase in the number of confirmed cases since the mid-1980s (Wee et al., 2008). Although there has been a national eradication program since the 1960s, an active surveillance program for brucellosis was not implemented before the 2000s (Yoo et al., 2009). The first human case in South Korea was officially reported in 2002, in a farm worker following the consumption of unpasteurized milk (Park et al., 2003). Thereafter, the number of human cases rapidly increased (Kim et al., 2006). In 2004, a new intensive brucellosis eradication program covering all dairy and beef cattle was launched. In South Korea, most human cases are related to not wearing protection, e.g. gloves, goggles and protective clothing, when in contact with suspected cattle or materials; but the consumption of raw milk and cheese is not common (Park et al., 2005).

Due to the zoonotic and economic aspects of this disease, count data are commonly collected for cattle and for human cases—typically through separate surveillance systems. Although monthly counts of human and cattle cases have been collected for several years in South Korea, the temporal relationship in the counts between species in the country has not been assessed. The relatively recent initiation of eradication programs in South Korea provided an opportunity to investigate the relationship between cattle and human count data obtained through independent systems. This type of time series data can be analyzed using an autoregressive integrated moving average (ARIMA) and Poisson [or negative binomial regression (NBR)] models. The different models have been used to analyze the time series data depending on their advantages and suitability. It was hypothesized that such a relationship could be quantified in a time series model and that such a model might have utility in predicting the impact of a reduction in cattle cases upon human case counts. The main objective of this study was to retrospectively develop an appropriate time series model of human and bovine brucellosis in South Korea using two methods and to compare their predictive capabilities.

## 2. Materials and methods

### 2.1. Data sources

National human and cattle population data were collected by the Korean Statistical Information Service (KOSIS) on a yearly and quarterly basis, respectively. Human and cattle cases were collected on a monthly basis by the Korea Centers for Disease Control and Prevention (KCDC) and the Animal Infectious Disease Data Management (AIMS), respectively. Both of these systems are operated by the South Korean government. Human case information was collected by passive surveillance. If humans were diagnosed with brucellosis at local or university hospitals, these cases were reported to the local public health authorities and captured into the central system of the KCDC. Cattle cases were reported by active and passive surveillance systems at the farm level. Dairy herds were tested six times a year using milk ring testing. If there were positive results, blood samples were collected and tested using the Rose-Bengal plate agglutination test. The beef cattle are tested twice a year on all farms that had more than 10 beef cattle using the Rose-Bengal plate agglutination test. In addition, slaughterhouse and pre-movement testing (between farms and markets) were mandatorily conducted. All the positive samples were retested using a serum agglutination test as a confirmatory test. Also, suspected cases were voluntarily reported to the authorities or veterinarians for laboratory testing. Laboratory testing of bovine samples was conducted at the National Veterinary Research and Quarantine Service, a World Organization for Animal Health (OIE) reference laboratory for brucellosis.

Since June 2004, intensive national surveillance and control measures (such as a brucellosis-free certificate system for sale or slaughter) have been conducted in all cattle. Therefore, we expected that the estimation of national cattle cases of brucellosis has become more accurate since 2004. Human and cattle case counts as recorded by the KCDC and AIMS, respectively were collected for the 6-year period from January 1, 2005, through December 31, 2010. From KOSIS, national human and cattle population data were obtained. All data sets were imported through Microsoft Excel 2007 (Redmond, WA, USA).

### 2.2. Time series analysis

The incidence rates for human and cattle were calculated on a monthly basis (cases/national total population) and reported per 100,000 population. In order to compute the incidence rates on a monthly basis, it was assumed that the national human and cattle population were constant on a yearly and quarterly basis, respectively, during this study period. Both crude incidence rates were used in the models. The human and cattle case datasets were divided into model construction (2005–2007) and validation (2008–2010) phases.

A time series ARIMAX model was first constructed, because it is statistically well developed and sophisticated model dealing with time series data. The ARIMAX model is an extension of the autoregressive integrated moving average (ARIMA) model, where external covariates may be added depending on cross-correlations between them and the response variable. Thus, an ARIMAX model was used because the cattle incidence rate should be included in the ARIMA model as an additional covariate. The common notation for the ARIMAX model is ARIMA($p$, $d$, $q$) – AR($p$), which is explained below. The stationarity of human incidence rate was assessed by plotting an autocorrelation function (ACF) (Diggle, 1990). Due to a lack of stationarity, the first order differencing was used with the purpose of stabilizing the response variable. Next, for assessing seasonality, a time sequence plot was used to identify any periodic fluctuations on a monthly basis. Once stationarity