Modeling to Predict Cases of Hantavirus Pulmonary Syndrome in Chile

Elaine O. Nsoesie1,2,3, Sumiko R. Mekaru1, Naren Ramakrishnan4, Madhav V. Marathe3,4, John S. Brownstein1,2,5

1 Children’s Hospital Informatics Program, Boston Children’s Hospital, Boston, Massachusetts, United States of America, 2 Department of Pediatrics, Harvard Medical School, Boston, Massachusetts, United States of America, 3 Network Dynamics and Simulation Science Laboratory, Virginia Bioinformatics Institute, Virginia Tech, Blacksburg, Virginia, United States of America, 4 Department of Computer Science, Virginia Tech, Blacksburg, Virginia, United States of America, 5 Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montreal, Canada

Abstract

**Background:** Hantavirus pulmonary syndrome (HPS) is a life threatening disease transmitted by the rodent *Oligoryzomys longicaudatus* in Chile. Hantavirus outbreaks are typically small and geographically confined. Several studies have estimated risk based on spatial and temporal distribution of cases in relation to climate and environmental variables, but few have considered climatological modeling of HPS incidence for monitoring and forecasting purposes.

**Methodology:** Monthly counts of confirmed HPS cases were obtained from the Chilean Ministry of Health for 2001–2012. There were an estimated 667 confirmed HPS cases. The data suggested a seasonal trend, which appeared to correlate with changes in climatological variables such as temperature, precipitation, and humidity. We considered several Auto Regressive Integrated Moving Average (ARIMA) time-series models and regression models with ARIMA errors with one or a combination of these climate variables as covariates. We adopted an information-theoretic approach to model ranking and selection. Data from 2001–2009 were used in fitting and data from January 2010 to December 2012 were used for one-step-ahead predictions.

**Results:** We focused on six models. In a baseline model, future HPS cases were forecasted from previous incidence; the other models included climate variables as covariates. The baseline model had a Corrected Akaike Information Criterion (AICc) of 444.98, and the top ranked model, which included precipitation, had an AICc of 437.62. Although the AICc of the top ranked model only provided a 1.65% improvement to the baseline AICc, the empirical support was 39 times stronger relative to the baseline model.

**Conclusions:** Instead of choosing a single model, we present a set of candidate models that can be used in modeling and forecasting confirmed HPS cases in Chile. The models can be improved by using data at the regional level and easily extended to other countries with seasonal incidence of HPS.

Citation: Nsoesie EO, Mekaru SR, Ramakrishnan N, Marathe MV, Brownstein JS (2014) Modeling to Predict Cases of Hantavirus Pulmonary Syndrome in Chile. PLoS Negl Trop Dis 8(4): e2779. doi:10.1371/journal.pntd.0002779

Editor: Daniel G. Bausch, Tulane School of Public Health and Tropical Medicine, United States of America

Received June 18, 2013; Accepted February 23, 2014; Published April 24, 2014

Copyright: © 2014 Nsoesie et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work is supported by the Intelligence Advanced Research Projects Activity (IARPA) via Department of Interior National Business Center (DoI/NBC) contract number D12PC000337. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: onelaine@vbi.vt.edu

Introduction

Reports indicate that illness from hantaviruses existed as early as the 1950s [1], although the virus was only officially recognized in the Americas in 1993 [1–3]. An outbreak of hantavirus in the Four Corners region of the United States led to investigation and isolation of the Sin Nombre virus (SNV). Thereafter, several other hantaviruses with the ability to cause severe human disease were discovered in rodents in the Americas [2], with SNV and Andes virus (ANDV) predominating in North and South America respectively [4]. Rodents are a natural reservoir for hantaviruses. Typically, different rodents carry different hantavirus strains and rodent species vary by geography. The deer mouse (*Peromyscus maniculatus*) is the primary reservoir for SNV [2,5–7], while *Oligoryzomys longicaudatus* commonly known as the long-tailed pygmy rice rat is the primary reservoir for the ANDV [3,4,8–10]. The long-tailed pygmy rice rat is common in rural Argentina and Chile [9,11].

Transmission of hantaviruses to humans occurs through contact with infected rodent saliva, excreta or inhalation of contaminated aerosol, and, rarely, through the bite of an infected rodent [12–14]. Outbreaks are typically small and geographically confined partly due to the lack of human-to-human transmission. Infections can progress to severe disease defined by geographical regions: hemorrhagic fever with renal syndrome (HFRS) in the Old World (Asia and Europe) and hantavirus pulmonary syndrome (or hantavirus cardiopulmonary syndrome) (HPS) in the New World (Americas) [4]. Symptoms of HPS include myositis, fever, breathing difficulties, and in some cases, vomiting, diarrhea, headaches, dizziness and kidney problems [13–15]. There are an estimated 300
Hantavirus pulmonary syndrome (HPS) is a severe disease present in Chile, Argentina and other countries in the Americas. Mortality rates for HPS can be as high as 60% for some outbreaks. Although hantavirus outbreaks tend to be small, the high death rate, unavailability of a vaccine, and occurrence of infections in rural regions where individuals are least likely to have appropriate healthcare make HPS an important public health issue in Chile and other countries. We present an approach for modeling and forecasting confirmed HPS cases in Chile.

Seasonal time series models that predict future cases based on previous cases appear reasonable. However, adding climate variables such as precipitation, which is thought to indirectly influence outbreaks of hantavirus slightly improves the model fit. To further improve the current models to make them more useful for public health preparedness/interventions, data at the regional level with reliable predictions several months into the future are needed.

Author Summary

Hantavirus pulmonary syndrome (HPS) is a severe disease present in Chile, Argentina and other countries in the Americas. Mortality rates for HPS can be as high as 60% for some outbreaks. Although hantavirus outbreaks tend to be small, the high death rate, unavailability of a vaccine, and occurrence of infections in rural regions where individuals are least likely to have appropriate healthcare make HPS an important public health issue in Chile and other countries. We present an approach for modeling and forecasting confirmed HPS cases in Chile. Seasonal time series models that predict future cases based on previous cases appear reasonable. However, adding climate variables such as precipitation, which is thought to indirectly influence outbreaks of hantavirus slightly improves the model fit. To further improve the current models to make them more useful for public health preparedness/interventions, data at the regional level with reliable predictions several months into the future are needed.

Materials and Methods

Data Sources

Official case count time series. Publicly available data on confirmed HPS cases were obtained from the Chilean Ministry of Health (http://epi.minsal.cl). We extracted the data from reports presented on the Ministry of Health (MOH) website. The data spanned January 2001 to December 2012 and was available at a monthly resolution (see Figure 1). At the time of this writing, the HPS map on the Chile’s MOH site showed a cumulative distribution of cases from 2007 to the week of May 26th, 2012. The case distribution from 2007–2012 by region was as follows: Los Ríos (105), Biobío (92), Maule (59), Los Lagos (47), Del Libertador B. O’Higgins (44), Araucanía (41) and Valparaíso (6) [28]. The number of reported HPS cases for this period were also compared and found to be equitable to the number of reported HPS cases extracted from HealthMap [29] - an automated online system for real-time disease outbreak monitoring and surveillance based on informal sources such as news reports.

Climate data. Climate data were obtained from the Global Historical Climatology Network (GHCN) of the US National Oceanic and Atmospheric Administration (NOAA) [30]. GHCN data was available for nine weather stations located across nine of the fifteen regions in Chile. We further restricted the data to stations located in regions with incidence of HPS. The GHCN data contained temperature and cumulative precipitation. Temperature was recorded in degree Celsius, and the average mean, maximum and minimum monthly values were presented. Precipitation was measured as total precipitation per month and presented at tenths of millimeter precision. Data on daily relative humidity and dew point were downloaded from NOAA’s National Climatic Data Center at a daily resolution. We chose 23 stations (see Figure 2) located in Los Lagos, Maule, Valparaíso, Biobío, Araucanía, Los Lagos and Aisen. We also downloaded the data from January 2001 to December 2012 to correspond with the time range of the reported HPS cases. In addition, monthly averages were estimated from the reported daily relative humidity and dew point. These climatological variables were chosen based on available data and previous studies on spatial spread of hantaviruses using similar explanatory variables [22,26].

Model Fitting

As stated, the aim of this study was to build models that estimate and forecast HPS activity in Chile, in addition to exploring the influence of climatological variables in the models. We selected six time series models after considering several Auto Regressive Integrated Moving Average (ARIMA) and regression with ARIMA error models. ARIMA models are useful for modeling lagged relationships typically present in periodically collected data [31]. Also, a lagged relationship between climate variables and cases has biologic plausibility as these variables may have lagged effect on rodent population and behavior and through that pathway a lagged effect on cases. ARIMA models have been used in modeling and forecasting infectious diseases such as influenza [32] and dengue [33]. We briefly describe these modeling approaches.

ARIMA models. ARIMA models belong to the Box-Jenkins approach to time-series modeling [34] and are represented as ARIMA(p,d,q), where p indicates the autoregressive (AR) order, d the differencing order and q the moving average (MA) order. ARIMA models assume a stationary time series, implying that the data vary around a constant mean and variance over time. Nonstationary time series variables can be made stationary by transformation or differencing. An ARIMA(p,d,q) model can be written as:

\[ y'_t = c + \phi_1 y'_{t-1} + \cdots + \phi_p y'_{t-p} - \theta_1 z_{t-1} - \cdots - \theta_q z_{t-q} + \epsilon_t \]  

where c is a constant, \( y'_t = y_t - y_{t-1} \) represents the differenced
seasonal AR(1) term, \((1-L)\) is the non-seasonal AR(1) term, \((1-L^s)\) is the seasonal difference term, \((1-\Phi_1 L^d)\) is the seasonal AR(1) term, \((1-\Theta_1 L^s)\) is the seasonal MA(1) term and \((1+\Theta_1 L^s)\) is the seasonal MA(1) term.

More information on ARIMA models can be found in several textbooks (for examples see [34] and [35]).

**Regression models with ARIMA errors.** Regression with ARIMA errors basically combines regression and ARIMA models. In general the model can be written as:

\[
y_t = \beta_1 x_{1,t} + \cdots + \beta_k x_{k,t} + n_t
\]

where \(n_t\) is assumed to follow an ARIMA model. For instance, if \(y_t\) and \(x_t\) are differenced, then a regression model with ARIMA(1,1,1) errors can be written as:

\[
y_t' = \beta_1 x_{1,t}' + \cdots + \beta_k x_{k,t}' + n_t'
\]

\[
(1-\phi_1 L)n_t' = (1+\theta_1 L)e_t
\]

where \(y_t'=y_{t-1}, x_{1,t}'=x_{1,t}-x_{1,t-1}, n_t'=n_t-n_{t-1}\) and \(e_t\) is a white noise series.

**Analysis.** Initially, we fitted a simple time series regression model to evaluate the potential impact of the climate variables on HPS cases. The residuals of each fitted model were evaluated and if the observed residuals did not correspond to white noise, we employed a three step modeling approach which involved selecting a candidate model; estimating the model and performing diagnostic tests and lastly, forecasting future observations.

Autocorrelation function (ACF) and partial autocorrelation function (PACF) plots were used to determine possible values for the AR and MA orders. AR and MA coefficients were estimated using conditional-sum-of-squares and maximum likelihood. Additionally, we also used the cross-correlation function (CCF) to identify lags of the climatological variables to include in the models. We only used lags up to 4 months due to the length of the series, \(y'_{t-p}\) are lagged values and \(z_t\) is a white noise process.

Model (1) can be written using the backshift operator \(L\), which represents lagging of the data by a single time period (e.g. one month). Similarly \(L^s\) implies the data is lagged by two time periods or \(y_{t-2}\). The first difference using the shift operator can be written as \((1-L)yt\) and \(d\)-order difference can be written as \((1-L)^d yt\). Using this notation, equation (1) can be written as:

\[
(1-\phi_1 L-\cdots -\phi_p L^p)(1-L)^d yt = c + (1+\theta_1 L+\cdots +\theta_q L^q)z_t
\]

where \(d\) represents the number of differences, \(q\) are the number of MA terms, \(p\) is the number of AR terms and \(\phi\) and \(\theta\) are coefficients.

We also fit seasonal ARIMA models, defined as ARIMA\((p,d,q)(P,D,Q)_s\), where \(P\) is the number of seasonal auto-regressive terms, \(D\) represents the number of seasonal difference, \(Q\) is the number of seasonal moving average terms and \(s\) is the number of periods in the season which was 12 in this study since HPS cases were reported at a monthly resolution. For example an ARIMA\((1,1,1)(1,1,1)_4\) without a constant can be presented as

\[
ARIMA(1,1,1)(1,1,1)_4
\]

which is the non-seasonal ARIMA(1,1,1) model without a constant. Where \(y_{t-1}\) is the non-seasonal MA(1) term and \(y_{t-1}L\) is the non-seasonal MA(1) term.

The first difference using the shift operator can be presented as \((1-L)yt\) and \(d\)-order difference can be written as \((1-L)^d yt\). Using this notation, equation (1) can be written as:

\[
(1-\phi_1 L-\cdots -\phi_p L^p)(1-L)^d yt = c + (1+\theta_1 L+\cdots +\theta_q L^q)z_t
\]

where \(d\) represents the number of differences, \(q\) are the number of MA terms, \(p\) is the number of AR terms and \(\phi\) and \(\theta\) are coefficients.

We also fit seasonal ARIMA models, defined as ARIMA\((p,d,q)(P,D,Q)_s\), where \(P\) is the number of seasonal auto-regressive terms, \(D\) represents the number of seasonal difference, \(Q\) is the number of seasonal moving average terms and \(s\) is the number of periods in the season which was 12 in this study since HPS cases were reported at a monthly resolution. For example an ARIMA\((1,1,1)(1,1,1)_4\) without a constant can be presented as

\[
ARIMA(1,1,1)(1,1,1)_4
\]

which is the non-seasonal ARIMA(1,1,1) model without a constant. Where \(y_{t-1}\) is the non-seasonal MA(1) term and \(y_{t-1}L\) is the non-seasonal MA(1) term.

More information on ARIMA models can be found in several textbooks (for examples see [34] and [35]).

**Regression models with ARIMA errors.** Regression with ARIMA errors basically combines regression and ARIMA models. In general the model can be written as:

\[
y_t = \beta_1 x_{1,t} + \cdots + \beta_k x_{k,t} + n_t \quad (4)
\]

where \(n_t\) is assumed to follow an ARIMA model. For instance, if \(y_t\) and \(x_t\) are differenced, then a regression model with ARIMA(1,1,1) errors can be written as:

\[
y_t' = \beta_1 x_{1,t}' + \cdots + \beta_k x_{k,t}' + n_t'
\]

\[
(1-\phi_1 L)n_t' = (1+\theta_1 L)e_t \quad (5)
\]

where \(y_t'=y_{t-1}, x_{1,t}'=x_{1,t}-x_{1,t-1}, n_t'=n_t-n_{t-1}\) and \(e_t\) is a white noise series.

**Analysis.** Initially, we fitted a simple time series regression model to evaluate the potential impact of the climate variables on HPS cases. The residuals of each fitted model were evaluated and if the observed residuals did not correspond to white noise, we employed a three step modeling approach which involved selecting a candidate model; estimating the model and performing diagnostic tests and lastly, forecasting future observations.

Autocorrelation function (ACF) and partial autocorrelation function (PACF) plots were used to determine possible values for the AR and MA orders. AR and MA coefficients were estimated using conditional-sum-of-squares and maximum likelihood. Additionally, we also used the cross-correlation function (CCF) to identify lags of the climatological variables to include in the models. We only used lags up to 4 months due to the length of the
time series. CCF was calculated between pre-whitened time series for each of the climate variables and the confirmed HPS cases. The variables were pre-whitened by first fitting an ARIMA model to the time-series of each of the climate variables. Next, the HPS case data were filtered using the estimated coefficients for each of the fitted models. Lastly, we assessed correlations between the residuals from each of the models fitted to the climate variables and the filtered confirmed HPS cases using the CCF. Pre-whitening is necessary when CCF values are affected by the time series structure of the covariates and existing common trends between the covariates and the response series over time [35]. Dew point was excluded from the modeling process because it did not have a significant CCF with the HPS cases. Models consisting of one or more of the remaining climate variables in addition to lagged climate variables were considered.

We adopted an information-theoretic approach for model and ranking. Model selection was based on a minimization of the Corrected Akaike Information Criterion (AICc) [36–38] and diagnostic checks. Diagnostic checks involved examining the residuals for each fitted model for autocorrelations and randomness. The difference ($\Delta$AICc) between the AICc of each selected model and the minimum AICc was also calculated. Furthermore, we estimated model likelihoods and Akaike weights $w_i$ as a measure of the strength of evidence in support of each of the candidate models given the data and the selected models. The weight was calculated as follows:

$$w_i \approx \exp \left( -\frac{\Delta \text{AIC}_c}{2} \right) \left/ \sum_{i=1}^{R} \exp \left( -\frac{\Delta \text{AIC}_c}{2} \right) \right)$$ [37]. Here, R represents the number of candidate models. Lastly, we estimated the evidence ratio between the model with minimum AICc and the model without climate variables by taking the ratio of the Akaike weights.

The monthly HPS case data were divided into two sets: one for the fitting process and the other for validation. Data from 2001–2009 were used in fitting. Data from January 2010 to December 2012 were used in validation; 106 data points were used in fitting and 36 for evaluation. The selected models were also used in one-step-ahead predictions of the evaluation points and prediction accuracy was assessed based on the coefficient of variation $R^2$ and the Root Mean Squared Error (RMSE). The modeling process was implemented using the TSA and forecast packages in R [39]. A sample code is given in the Supplementary Information S1.

**Results**

From 2001–2012, there were an estimated 667 cases of HPS with an average of 56 cases per year. HPS cases were relatively consistent among years with seasonal peaks between January and April as shown in Figure 1. The epidemic peaks appeared to be preceded by peaks in mean temperature. In addition, increases in precipitation and relative humidity appeared to correlate with troughs in confirmed HPS cases (see Figure 3). Cross-correlations between pre-whitened time series of climate variables and the confirmed HPS case data were significant at different lags, mostly between lags 1 and 4. The cross-correlations of variables used in the modeling process are shown in Table 1.

Five models were selected in addition to the model based solely on previous HPS cases (baseline model). The models are presented based on AICc rankings in Table 2. The model with precipitation with lags at one and four months was the most parsimonious with...
Figure 3. Confirmed HPS cases in Chile and climate variables from 2001–2012. Abbreviations: Temp = Temperature, Rhx = Relative Humidity and Prcp = Precipitation. The data is at a monthly time scale.
doi:10.1371/journal.pntd.0002779.g003
Table 1. Cross-correlations between climate time series and reported HPS cases.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Lag in months</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tmin</td>
<td></td>
<td>-0.017</td>
<td>-0.034</td>
<td>-0.055</td>
<td>0.012</td>
<td>0.063</td>
<td>0.032</td>
<td>0.082</td>
<td>0.035</td>
<td>0.014</td>
<td>0.005</td>
<td></td>
</tr>
<tr>
<td>Tmax</td>
<td></td>
<td>0.012</td>
<td>0.032</td>
<td>0.055</td>
<td>0.012</td>
<td>0.063</td>
<td>0.032</td>
<td>0.082</td>
<td>0.035</td>
<td>0.014</td>
<td>0.005</td>
<td></td>
</tr>
<tr>
<td>Prcp</td>
<td></td>
<td>-0.007</td>
<td>-0.017</td>
<td>-0.034</td>
<td>0.012</td>
<td>0.063</td>
<td>0.032</td>
<td>0.082</td>
<td>0.035</td>
<td>0.014</td>
<td>0.005</td>
<td></td>
</tr>
<tr>
<td>Rhx</td>
<td></td>
<td>0.003</td>
<td>0.006</td>
<td>0.011</td>
<td>0.016</td>
<td>0.025</td>
<td>0.037</td>
<td>0.021</td>
<td>0.019</td>
<td>0.010</td>
<td>0.002</td>
<td></td>
</tr>
</tbody>
</table>

Tmin represents monthly average minimum temperature, Tmax is the monthly average max temperature, Prcp is cumulative monthly precipitation and Rhx is relative humidity. Each variable was fitted to an ARIMA model and the HPS case time series data was filtered using the estimated coefficients from each of the fitted models. Correlations were then assessed between the residuals from each of the models fitted to the climate variables and the filtered confirmed HPS cases using the cross-correlation function. Significant correlations are in bold.

Discussion

In this study, we present several time series models for HPS activity with the inclusion and exclusion of climatological covariates. Univariate models with a seasonal component appear to capture the trend observed in the data and also have a low prediction error suggesting that HPS cases could be predicted one-month ahead. The inclusion of climatological variables slightly improves the model fit and prediction accuracy in some cases. Since HPS cases are typically low, slight deviations in fit and prediction could be significant. The model based on previous observations of HPS cases and precipitation is ranked highest based on the AICc. Increased rainfall may increase the availability of food resources for rodents, which may lead to higher reproduction. Growth in rodent populations may lead to competition for food resources, which may also lead to greater dispersal thereby increasing the likelihood of contact between rodents and humans [21]. As this process takes place overtime, it may partially explain some of the significant correlations observed between lags of the pre-whitened precipitation variable and the confirmed HPS case time series data.

Though some studies have evaluated climate in relation to rodent population dynamics, to our knowledge, there are no published studies on forecasting HPS activity using climate variables for Chile. A study on forecasting HFRRS incidence in China used a similar time series modeling approach [31]. However, the authors did not consider the inclusion of climate variables in their model. The availability of climate data in near real-time makes a reasonable addition to the modeling and forecasting of HPS cases.

Climate most likely influences rodent populations based on a combination of weather variables [7,20,21,40]. However, there is a lack of in-depth understanding on how climate change would influence rodent host populations and disease outbreaks. Recent changes in rodent populations and hantavirus infections have been connected with climate change in Europe, though the effects appear to differ from one location to another [3]. However, the natural history and hosts dynamics of hantaviruses suggest that climate change could result in increased incidence in humans, which might also be attributable to changes in human behavior [3,8,40]. In addition to the possible elimination of reservoir hosts, increases in the frequency and intensity of extreme climatic events such as droughts and floods, could lead to alteration of hantavirus species composition [8,41]. A more in-depth understanding of both reservoir rodent population dynamics and disease outbreaks in relation to climate is therefore important for better understanding and modeling of ANDV infections in rodents.
### Table 2. The models are ranked based on AICc and presented by rank order.

<table>
<thead>
<tr>
<th>Model Errors</th>
<th>Fit</th>
<th>Prediction</th>
<th>AR</th>
<th>MA</th>
<th>Environmental Variables</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AICc</td>
<td>Δ AICc_i</td>
<td>w_i</td>
<td>RMSE</td>
<td>R^2</td>
</tr>
<tr>
<td>ARIMA(1,0,0)(0,1,1)_{12}</td>
<td>437.62</td>
<td>0</td>
<td>0.630</td>
<td>2.557</td>
<td>0.559</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ARIMA(2,0,0)(0,1,1)_{12}</td>
<td>440.30</td>
<td>2.68</td>
<td>0.165</td>
<td>2.774</td>
<td>0.551</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ARIMA(2,0,1)(0,1,1)_{12}</td>
<td>440.90</td>
<td>3.28</td>
<td>0.122</td>
<td>2.542</td>
<td>0.558</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ARIMA(2,0,0)(0,1,1)_{12}</td>
<td>443.31</td>
<td>5.69</td>
<td>0.037</td>
<td>2.421</td>
<td>0.594</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ARIMA(1,0,0)(0,1,1)_{12}</td>
<td>444.98</td>
<td>7.36</td>
<td>0.016</td>
<td>2.560</td>
<td>0.555</td>
</tr>
</tbody>
</table>

The errors terms for each of the regression with autoregressive integrated moving average errors models is presented as ARIMA(p,d,q)(P,D,Q)s where p indicates the autoregressive (AR) order, d the differencing order, q the moving average (MA) order, and P, D, Q are the seasonal equivalents, and s is the seasonal period. The Akaike ΔAICc is the difference between each AICc and the minimum AICc. The weight w_i represents the probability of each model given the data and the other candidate models in the set. Prediction accuracy is evaluated based on the R^2 and RMSE. In addition, coefficients for each of the terms in the model in addition to the error terms are presented. Abbreviations: Rhx is relative humidity, Tmin is the minimum temperature, and Prcp is precipitation. The baseline model, which is based solely on previous confirmed cases of HPS is **bold** in the table.

doi:10.1371/journal.pntd.0002779.t002
Modeling to Predict HPS Cases in Chile

Model based on previous observations of HPS cases

Model based on previous HPS cases and monthly precipitation

Cases

Jan-01  Jan-02  Jan-03  Jan-04  Jan-05  Jan-06  Jan-07  Jan-08  Jan-09  Jan-10  Jan-11  Jan-12

Cases

Jan-01  Jan-02  Jan-03  Jan-04  Jan-05  Jan-06  Jan-07  Jan-08  Jan-09  Jan-10  Jan-11  Jan-12
The temporal and spatial resolutions of the available data are limitations of this study. Data at the daily or weekly resolution would be useful. In addition, averaging climate variables across different regions with different levels of HPS cases results in loss of information. Regional modeling and forecasts would be beneficial since weather varies across regions. This loss of information could have affected the contribution of the climate variables to the various models. The models also fail to account for several other factors that could possibly influence rodent populations and human and subsequently hantavirus infections. Other climate variables apart from humidity, precipitation, and temperature might also interact to influence observed cases. In addition, climate data is also available only for some of the regions with high HPS incidence and the locations of weather stations might not be the same as locations with confirmed cases. Furthermore, HPS cases typically occur in rural areas suggesting that not all cases might be reported and mild HPS cases might be undiagnosed. Riquelme et al. [9] suggests using epidemiological questionnaires to improve monitoring and diagnosis of HPS. There is also a likely delay in diagnosis since the average incubation period for hantavirus infections is between 13–17 days [1,13] possibly meaning some cases could be classified by a month. Lastly, the confidence intervals for the presented models include negative values suggesting that a Poisson time series model might be better suited for this data. However, note, that the predicted cases do not fall below zero, indicating that the models presented could be useful for case count predictions.

Although the proposed models have limitations, they provide a framework for monitoring and forecasting HPS cases. HPS is a serious disease with a high fatality ratio. Modeling and forecasting of expected HPS cases could be extremely useful for public health education and control of incidence. The possibility of person-to-person transmission [10,42] and potential vulnerability of urban communities makes such a modeling approach relevant. In addition, although several environmental management practices and educational campaigns have been launched to limit activities that contribute to the transmission of hantaviruses, HPS continues to be endemic in Chile [12]. Vaccines are currently unavailable, therefore prevention of hantavirus infection involves elimination of possible habitats for rodents, proper garbage disposal practices, and using protective devices (such as gloves, goggles and respirators) when cleaning poorly ventilated structures such as cabins and ensuring proper maintenance in and around household structures. Most hantavirus infections are thought to result from exposure to aerosolized hantavirus. Modeling and forecasting future HPS cases could be useful for the development and implementation of preventive public health measures. Although the models in this study are developed for Chile, methods and analysis can be extended to other countries with HPS.

These models must be updated with the most recent data regularly since changes in various parameters can affect predicted outcomes. Despite the noted limitations and requirements, these models suggest that reliable forecast of HPS cases in Chile is feasible. Solely using past cases to predict future case is possible due to the seasonal nature of the peaks. However, the inclusion of climate variables at an appropriate resolution can significantly improve the prediction of incidence. Although hantavirus outbreaks tend to be small and geographically confined, the high fatality rate, unavailability of a vaccine and occurrence of infections in rural regions where individuals are least likely to have appropriate healthcare makes HPS forecasting an important public health challenge in Chile and several other countries.

Supporting Information

Supporting Information S1 Sample R code for fitting models with ARIMA errors.

(TXT)

Acknowledgments

We would like to thank Michael Close and Michael Johansson for aiding with data extraction and model suggestions.

The US Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright annotation thereon. Disclaimer: The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of IARPA, Do/DRC or the US Government.

Author Contributions

Conceived and designed the experiments: EON SRM. Performed the experiments: EON. Analyzed the data: EON. Contributed reagents/materials/analysis tools: NR MVM JSB. Wrote the paper: EON SRM NR MVM JSB.

References


