Predicting the Incidence of Brucellosis in Western Iran using Markov switching model

Maryam Mohammadian-Khoshnoud
Hamadan University of Medical Sciences Medical School
Majid Sadeghifar (sadeghifar@basu.ac.ir)
Bu Ali Sina University https://orcid.org/0000-0001-6147-6051
Zahra Cheraghi
Hamadan University of Medical Sciences Medical School
Zahra Hosseinkhani
Qazvin University of Medical Sciences

Research note

Keywords: Brucellosis, zoonosis, climatic factors

Posted Date: July 22nd, 2020

DOI: https://doi.org/10.21203/rs.3.rs-42636/v1

Version of Record: A version of this preprint was published on March 1st, 2021. See the published version at https://doi.org/10.1186/s13104-020-05415-5.

Abstract

Objective: Brucellosis is a zoonosis almost chronic disease. Brucellosis bacteria can remain in the environment for a long time. Thus, climate irregularities could pave the way for the survival of the bacterium Brucellosis. The aim of this study is to investigate the effect of climatic factors as well as predicting the incidence of Brucellosis in Qazvin province using the Markov switching model. This study is a secondary study of data collected from 2010 to 2019 in Qazvin province. The data include Brucellosis cases and climatic parameters. Two state Markov switching model with time lags of zero, one and two was fitted to the data. The Bayesian information criterion was used to evaluate the models.

Results: According to the Bayesian information criterion, the two-state Markov switching model with a one-month lag is a suitable model. The month, the average wind speed, the minimum temperature have a positive effect on the number of Brucellosis, the age and rainfall have a negative effect. The results show that the probability of an outbreak for the third month of 2019 is 0.30%.

Introduction

Brucellosis is a zoonosis almost chronic disease which is transmitted by direct or indirect contact with infected animals or products(1). Brucella melitensis is the most common and acute pathogen that can cause various symptoms such as sudden tremors, general body aches (2, 3, 4). This disease causes many economic and health problems for communities and these problems are more severe in communities where animal husbandry is the main occupation. The disease has disappeared in many developed countries such as Australia, Japan, but in the Mediterranean, Middle East, and parts of Asia is still as a major public health problem(5). Despite a good health care system in Iran, Brucellosis is still an important endemic disease. Iran ranks fourth in the world in terms of the incidence of Brucellosis(5–7).

In Iran, the incidence of Brucellosis is reported to be between 50 and 500 per 100,000 people, often of the B.melitensis type (8) and is known as a local disease. Brucellosis is more common in men 25 to 29 years of age, in the western provinces, and in the spring months(9). The incidence rate of Brucellosis in Qazvin province is 27.43 per 100,000 during 2010–2019(10).

Climate is one of the most important factors affecting human health, especially in the case of infectious diseases(11). Optimal climate conditions are critical to the survival, reproduction, distribution, and transmission of pathogen and host(12).

Despite the significant effect of climatic factors on brucellosis, these factors have been overlooked in most studies. The primary purpose of this study was to investigate the effect of climatic factors on the incidence of Brucellosis. The second purpose of this study is to use Markov switching model (MSM), which are considered as nonlinear time series models(13). A number of researchers have used the MSM in disease prediction(14) but so far the performance of this method has not been evaluated in Brucellosis data. In this study, 8-years Brucellosis time series data were used in Qazvin province. Thus, this study will be conducted with the two objectives of examining the factors related to Brucellosis and predicting the incidence of Brucellosis in Qazvin province during 2010 to 2019 using the Markov switching model.

Method

This study is a secondary study of data collected from April 2010 to March 2019 in Qazvin province, which is extracted from the database of the Qazvin University of Medical Sciences and the meteorological system of the province(10). No individual data were used. The available information was cumulative without mentioning first and last name and other personal information.

The response variable is the number of Brucellosis cases. Month, rural ratio, age, men Ratio, the ratio of contact with livestock, Non-pasteurized dairy, average monthly temperature, total monthly rainfall, average wind speed, maximum monthly temperature, minimum monthly temperature and wind speed are considered as explanatory variables.

MSM including time series $\{Y_t\}_{t=1,\ldots,T}$ and a sequence of related variables $x_1,\ldots,x_T$ is introduced with the relation between x_t and Y_t as follows:

$$Y_t = \boldsymbol{f}^{(s_t)}(\boldsymbol{x_t}) + \boldsymbol{\sigma_{st}\epsilon_t}$$

Where $\boldsymbol{\epsilon_t}$ has a standard normal distribution and S_t is the state at time t of a non-observable N state Markov chain. In two-state models, if we define state 1 as the disease outbreak period and zero state as the non-outbreak period, the probability of an outbreak in period *t + 1* can be as follows(14):

$$P(s_{t+1} = 0) = \frac{(1 - p_{11})}{(2 - p_{00} - p_{11})}$$

$$P(s_{t+1} = 1) = \frac{(1 - p_{00})}{(2 - p_{00} - p_{11})}$$

A two-state MSM with switching all effects is considered. In order to perform the required analyzes, SPSS software version 26 and R version 3.6.3 were used. The msmFit function in the MSwM package was used to fit the MSM using EM algorithm(15).

Results

From 2010 to 2018, 3194 people were infected with Brucellosis. Of these, the highest incidence related to the year 2015 with 512 (16%), and the lowest incidence related to the year 2010 with 192 (6%). Among the 4 seasons of the year, the summer season with 961 (30.1%) has the highest number of infected and in the next rank of winter with 805 (25.2%). The highest number of patients is related to temperatures 26 c^0 (6.6%), 25 c^0 (6.4%). Comparing the average monthly rainfall, the highest number of infected people is related to zero rainfall of 444 cases (13.9%) and the lowest number of patients is related to the average monthly rainfall of 17.8 (0.2%) and 64.5 (0.2%). The lowest number of patients is related to the average wind speed of 0.7 with 6 cases (0.2%).

Fitting MSM

To analyze, the MSM was fitted with two and three states and both models were fitted with 0, -1, and − 1 lags for climate variables. Temporal lag is defined as the time interval between climatic characteristics and the incidence of Brucellosis. Based on a comparison between 6 fitted models, based on Bayesian information criterion, two-state MSM with a time lag of one month is suitable. For this reason, only this model is offered to provide more results.

Age, month, rural ratio, men ratio, Non-pasteurized dairy, average monthly temperature, total monthly rainfall, average wind speed, maximum monthly temperature, Minimum monthly temperature, average wind speed, ratio of contact with livestock were recognized as significant variables (Table 1).

Table 1
The fit of the two-state markov switching model with time lags of 0, -1, -1

Variables(state)	MSM with two states								
	Lag 0			Lag – 1			Lag – 2		
	B	SE	p-value	B	SE	p-value	B	SE	p-value
Intercept(1)	3.274	3.073	0.287	9.981	11.531	0.387	33.426	11.879	0.005
Intercept(2)	10.010	11.628	0.389	3.254	3.047	0.286	-2.162	14.711	0.883
Age(1)	-0.152	0.016	0.000	-0.116	0.118	0.326	-0.250	0.059	0.000
Age(2)	-0.116	0.118	0.327	-0.152	0.016	0.000	0.499	0.224	0.026
Month (1)	1.618	0.108	0.000	0.817	0.371	0.028	1.164	0.408	0.004
Month (2)	0.816	0.371	0.028	1.618	0.108	0.000	-0.045	0.585	0.939
Rural Ratio(1)	3.172	0.113	0.000	-0.287	0.389	0.461	2.666	0.513	0.000
Rural Ratio(2)	-0.287	0.389	0.461	3.173	0.113	0.000	0.164	0.365	0.653
Men Ratio(1)	2.920	0.079	0.000	-3.009	0.665	0.000	-3.013	0.608	0.000
Men Ratio (2)	-3.012	0.667	0.000	2.928	0.100	0.000	1.412	1.517	0.352
Non-pasteurized dairy (1)	-0.570	0.070	0.000	0.569	0.229	0.013	1.056	0.296	0.000
Non-pasteurized dairy (2)	0.570	0.229	0.0127	-0.571	0.069	0.000	-0.273	0.230	0.235
Average monthly temperature (1)	-4.223	0.135	0.000	0.241	0.820	0.769	0.935	0.859	0.276
Average monthly temperature (2)	0.239	0.852	0.779	-4.229	0.142	0.000	-2.546	1.017	0.012
Total monthly rainfall (1)	-0.038	0.015	0.012	-0.034	0.041	0.405	-0.242	0.048	0.000
Total monthly rainfall (2)	-0.034	0.041	0.405	-0.038	0.015	0.011	0.009	0.053	0.850
Average wind speed (1)	1.509	0.214	0.000	19.024	2.848	0.000	18.097	3.285	0.000
Average wind speed (2)	19.017	2.869	0.000	1.533	0.185	0.000	5.163	3.336	0.122
Maximum monthly temperature (1)	1.717	0.119	0.000	-0.177	0.567	0.754	-1.377	0.597	0.021

Variables(state)	MSM with two states								
	Lag 0			Lag – 1			Lag – 2		
	B	SE	p-value	B	SE	p-value	B	SE	p-value
Maximum monthly temperature (2)	-0.176	0.585	0.763	1.718	0.118	0.000	0.722	0.629	0.251
Minimum monthly temperature (1)	3.203	0.096	0.000	0.231	0.408	0.572	0.345	0.403	0.393
Minimum monthly temperature (2)	0.231	0.417	0.579	3.207	0.101	0.000	2.289	0.599	0.000
Wind speed (1)	0.944	0.072	0.000	-0.698	0.301	0.020	-0.433	0.250	0.083
Wind speed (2)	-0.699	0.302	0.020	0.944	0.072	0.000	0.469	0.288	0.103
Ratio of contact with livestock (1)	-0.387	0.059	0.000	0.904	0.280	0.001	0.283	0.221	0.200
Ratio of contact with livestock (2)	0.904	0.280	0.001	-0.387	0.059	0.000	1.212	0.561	0.031

Autocorrelation and partial autocorrelation of residual and squared residual for model is confirmed lack of autocorrelation in the residual and the model seems to fit logically and there's no serial dependency on the residual.

Figure 2 shows the smoothed and filtered probabilities for state of one and two. Smoothed probabilities are used to determine peaks and depressions and 0.5 is determined as the cut-off value for zero and one states. The filtered probabilities are calculated using the first observation up to t and the smoothed probabilities are calculated using the total observations.

The Q-Q plot shows where the normality hypothesis is questionable for series. Transition probability matrix in MSM with a time lag of one month as follows:

$$\begin{bmatrix} P_{00} & P_{01} \\ P_{10} & P_{11} \end{bmatrix} = \begin{bmatrix} P_{00} & 1 - P_{00} \\ 1 - P_{11} & P_{11} \end{bmatrix} = \begin{bmatrix} 0.72 & 0.28 \\ 0.65 & 0.35 \end{bmatrix}$$

Where the probability of non-outbreak state in both t and t + 1 periods is 0.72, the probability of changing series from non-outbreak state in period t to outbreak state in period t + 1 is 0.28 and the probability of the series changing from the outbreak state in period t to non- outbreak state in period t + 1 is 0.65. When we are in a non-outbreak state (0.72), the process tends to stay the same state and the process is transferred to the outbreak state with a probability of 0.28.

The probability of an outbreak in t + 1 is as follows:

$$P\left[S_{t+1} = 0\right] = \frac{1 - 0.35}{2 - 0.72 - 0.35} = 0.70$$

$$P\left[S_{t+1} = 1\right] = 0.30$$

Since the data is up to the second month of 2019, the probability of an outbreak for the third month of 2019 (one month later) is very low and is equal to 0.30%.

The biggest difference between the coefficients of the variables in two states is related to the average wind speed. Therefore, the average wind speed is the most important factor in incidence Brucellosis.

In MSM, month, average wind speed and minimum temperature coefficients are positive which indicate a positive effect on the number of Brucellosis. In this model, the age and total monthly rainfall coefficients are negative, indicating a negative effect on the number of Brucellosis.

Discussion

The highest incidence of brucellosis is related to 1994 with 512 cases (16%). The average wind speed in 8 years was 1.89. The incidence is the highest at zero total monthly rainfall. Total monthly rainfall was 0 at 444 days (13.9%), which includes most days of study. The minimum temperature was − 1 and the maximum temperature was 30 degrees centigrade. The mean age of the patients was about 38 years.

There is no clear pattern in the number of cases of Brucellosis in the 8 years studied and fluctuations in the incidence of this disease can be seen with three peaks 2015 December, July 2015, June 2014. These results are inconsistent with the results of Lee's study. In the Lee study, the incidence of human Brucellosis in South Korea peaked in September 2006 and has dropped dramatically since then which indicates effective eradication (16). In Rafiemanesh’s study, the incidence of Brucellosis decreased from 2007 to 2016 which indicates an increase in the coverage of prevention programs, especially livestock vaccination(17). These results are inconsistent with the results of the present study.

During the 8 years, in the 2010 year, the lowest number of cases has been reported, followed by an upward trend until 2011. The reason for the rising trend of the disease from 2010 to 2011 may be related to the improvement of the data registration in the country's health system. This result is consistent with Hashtarkhani’s study(18). From 2014 to 2017, there is an upward trend in the number of Brucellosis cases which is inconsistent with Hashtarkhani's study. In the study of Hashtarkhani after 1990, we see a decreasing trend in the incidence of the disease(18). The results of the study show that there are the highest number of infected people in summer and winter seasons. These results are consistent with the results of the Tapak’s study. The results of the Tapak’s study show that hot summers and cold winters make the disease less common while climate moderation in these seasons exacerbates the disease(19). Therefore, the temperate climate of these seasons in Qazvin province increases the number of patients with Brucellosis. Model fit results indicate the negative effect of age and total monthly rainfall on the number of Brucellosis. This result is consistent with the results of Entezari's study. The results of the Entezari's study indicate a negative relationship between rainfall and brucellosis. In fact, as the rainfall decreases, the number of infected people increases(20).

Comparing the coefficients of explanatory variables in two state, the average wind speed is the most important factor in incidence Brucellosis. This result is consistent with the results of of Ahmadkhani's study. The results of Ahmadkhani's study indicate a positive correlation between wind speed, temperature, greenness and incidence of Brucellosis(21). The results of the present study are inconsistent with those of Tapak. According to the Tapak's study, the wind at high speeds reduces the disease. This is because the bacterium has a shorter lifespan in the air.

The mean age of the patients was about 38 years. This result has been confirmed in other studies(22, 23). The results of this study indicate that P_{00} and P_{11} are larger than P_{10} and P_{01}, respectively. In other words, states do not tend to change. That is, when we are in a non-outbreak state, the process tends to remain the same state. Also, P_{10} is higher than P_{01}, which indicates that prolonged periods of non-outbreak lead to a reduction in the probability of outbreak during the year.

As the country's health progresses in many areas, the incidence of Brucellosis is expected to decline. However, the results of the present study indicate a sharp increase in the disease between 2014 and 2017, which requires a lot of health attention in Qazvin province. The necessary learning for high-risk age and occupational groups should be on the agenda of the region's health institutions. Not consuming unpasteurized dairy products and not having contact with suspicious animals and cooperating with livestock vaccination are among the preventive health behaviors to reduce the number of new cases of the disease.

Conclusion

The MSM can be used to detect factors related to the incidence of brucellosis as well as to predict the incidence of brucellosis. Most climatic parameters were effective in incidence the disease, and the most influential factor was the average wind speed. The probability of disease outbreak in the third month of 2019 was predicted to be 0.30%.

Limitation

One of the limitations of this study is the limited period of the time series data and lack of daily information. Another limitation is the lack of comparison between different time series models.

Abbreviations

MSM
Markov switching model

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and material

Not applicable.

Competing interests

We declare that we have no competing interest.

Funding

This study (980217943, IR.UMSHA.REC.1398.085) was funded by the Vice-Chancellor of Research and Technology of Hamadan University of Medical Sciences. The funders had no role in study design, data collection, and analysis, decision to publish, or preparation of the manuscript.

Acknowledgements

We would like to thank the staffs of the Department of Zoonotic Diseases at the Center for Communicable Disease Control, Ministry of Health and Medical Education of Iran (MOHME). Moreover, we thank the Vice-Chancellor of Research and Technology of Hamadan University of Medical Sciences.

Contributions

Conceptualization: Maryam Mohammadian-Khoshnoud, Majid Sadeghifar, Zahra Cheraghi, Zahra Hosseinkhani.

Data curation: Zahra Cheraghi, Zahra Hosseinkhani.

Formal analysis: Maryam Mohammadian-Khoshnoud.

Funding acquisition: Zahra Cheraghi.

Methodology: Maryam Mohammadian-Khoshnoud, Majid Sadeghifar.

Writing – original draft: Maryam Mohammadian-Khoshnoud, Majid Sadeghifar.

Writing – review & editing: Maryam Mohammadian-Khoshnoud, Majid Sadeghifar, Zahra Cheraghi, Zahra Hosseinkhani.

References

1. Corbel M. Brucellosis in humans and animals2006.
2. Franco M, Mulder M, Gilman R, Smits H. Human brucellosis. The Lancet infectious diseases. 2007;7(12):775–86.
3. Silva T, Costa E, Paixão T, Tsolis R, Santos R. Laboratory animal models for brucellosis research. Journal of Biomedicine Biotechnology. 2011;2011:1–9.
4. Zeinali M, Shirzadi M, Sharifian J. National guideline for Brucellosis control. Tehran: Ministry of Health and Medical Education; 2009. pp. 10–7.
5. Pappas G, Papadimitriou P, Akritidis N, Christou L, Tsianos E. The new global map of human brucellosis. The Lancet infectious diseases. 2006;6(2):91–9.
6. Haran M, Agarwal A, Kupfer Y, Seneviratne C, Chawla K, Tessler S. Brucellosis presenting as septic shock. Case Reports. 2011;2011.
7. Ramin B, MacPherson P. Human brucellosis. Bmj. 2010;341:c4545.
8. Farahani S, Shah Mohamadi S, Navidi I, Sofian M. An investigation of the epidemiology of brucellosis in Arak City, Iran,(2001–2010). Journal of Arak University of Medical Sciences. 2012;14(7):49–54.
9. Bagheri H, Tapak L, Karami M, Amiri B, Cherghi Z. Epidemiological features of human brucellosis in iran (2011–2018) and prediction of brucellosis with data-mining models. Journal of Research in Health Sciences. 2019;19(4):e00462.
10. Bagheri H, Tapak L, Karami M, Hosseinkhani Z, Najari H, Karimi S, et al. Forecasting the monthly incidence rate of brucellosis in west of Iran using time series and data mining from 2010 to 2019. PloS one. 2020;15(5):e0232910.
11. Faramarzi H, Nasiri M, Khosravi M, Keshavarzi A, Ardakani A. Potential Effects of Climatic Parameters on Human Brucellosis in Fars Province, Iran, during 2009–2015. Iranian Journal of Medical Sciences. 2019;44(6):465.
12. Wu X, Lu Y, Zhou S, Chen L, Xu B. Impact of climate change on human infectious diseases: Empirical evidence and human adaptation. Environment international. 2016;86:14–23.
13. Hamilton J. A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle. Econometrica: Journal of the Econometric Society. 1989;57(2):357 – 84.
14. Ansari H, Mansournia M, Izadi S, Zeinali M, Mahmoodi M, Holakouie-Naieni K. Predicting CCHF incidence and its related factors using time-series analysis in the southeast of Iran: comparison of SARIMA and Markov switching models. Epidemiology Infection. 2015;143(4):839–50.
15. Sanchez-Espigares J. A L-M. fitting markov switching models. 2018.
16. Lee H, Her M, Levine M, Moore G. Time series analysis of human and bovine brucellosis in South Korea from 2005 to 2010. Prev Vet Med. 2013;110(2):190–7.

17. Rafiemanesh H, Alimohamadi Y, Aghdam S, Safarzadeh A, Shokri A, Zemestani A. Time series and trend analysis of brucellosis in Oskou county, East Azerbaijan: 2007–2016. Health promotion perspectives. 2019;9(4):285.
18. Hashtarkhani S, Akbari M, Jarahi L, Etminani K. Epidemiological characteristics and trend of incidence of human brucellosis in Razavi Khorasan province. Medical journal of Mashhad university of medical sciences. 2015;58(9):531–8.
19. Tapak L, Shirmohammadi-Khorram N, Hamidi O, Maryanaji Z. Predicting the frequency of human brucellosis using climatic indices by three data mining techniques of radial basis function, multilayer perceptron and nearest Neighbor: A comparative study. Iranian Journal of Epidemiology. 2018;14(2):153–65.
20. Entezari M, Moradpour S, Amiri M. Spatial distribution and the impact of geographical factors on brucellosis in Chaharmahal and Bakhtiari Province, Iran. International Journal of Epidemiologic Research. 2016;3(2):98–105.
21. Ahmadkhani M, Alesheikh A. Space-time analysis of human brucellosis considering environmental factors in Iran. Asian Pacific Journal of Tropical Disease. 2017;7(5):257–65.
22. Al-Tawfiq J, AbuKhamsin A. A 24-year study of the epidemiology of human brucellosis in a health-care system in Eastern Saudi Arabia. J Infect Public Health. 2009;2(2):81–5.
23. Buzgan T, Karahocagil M, Irmak H, Baran A, Karsen H, Evirgen O, et al. Clinical manifestations and complications in 1028 cases of brucellosis: a retrospective evaluation and review of the literature. International journal of infectious diseases. 2010;14(6):e469-78.

Figures

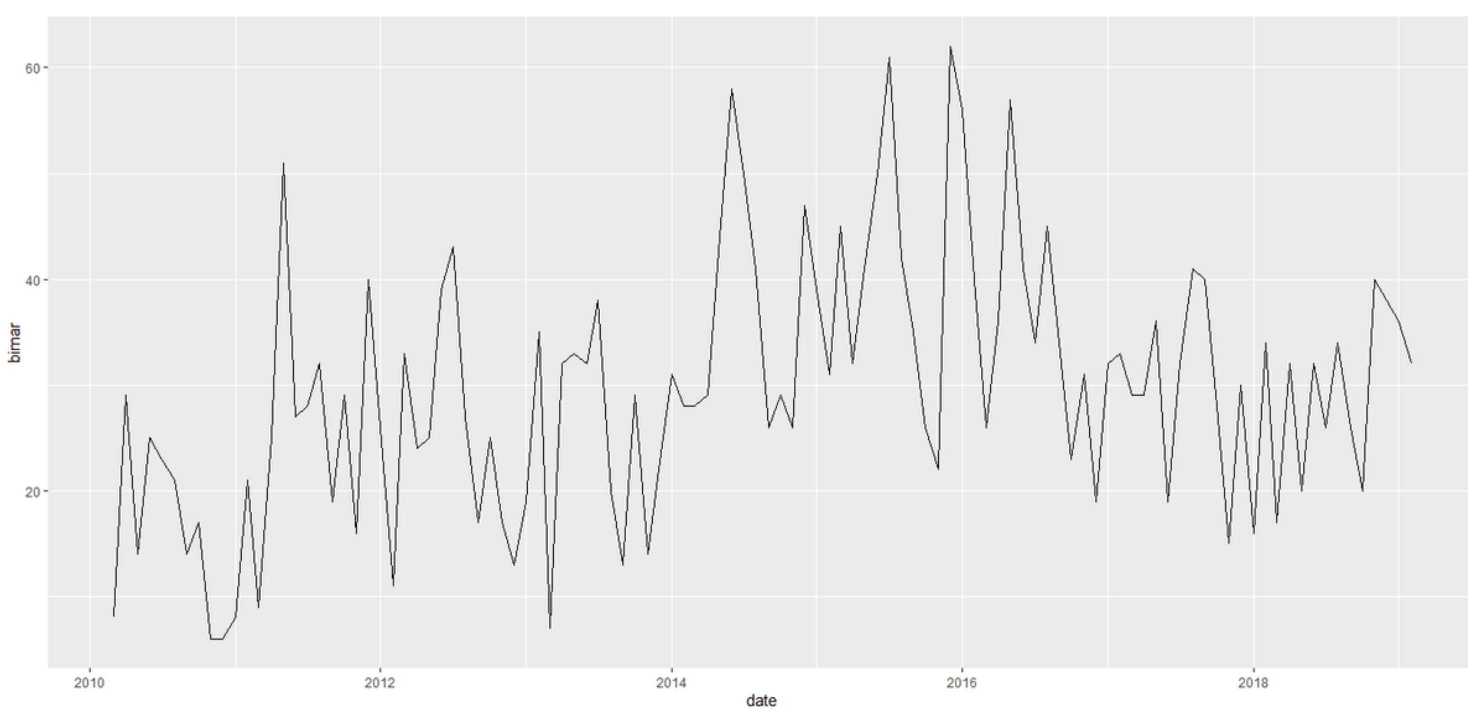

Figure 1

The monthly trend chart for the number of Brucellosis cases from 2010 to 2018

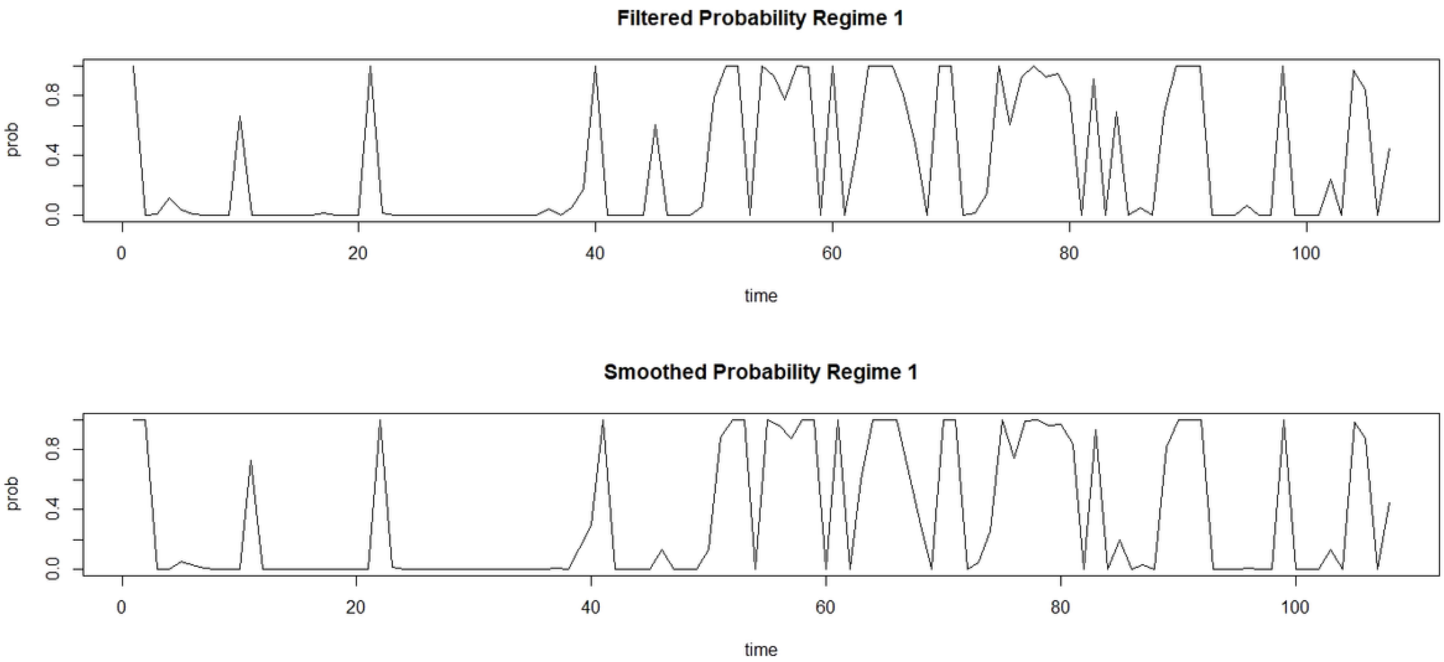

Figure 2

Probability of smoothed and filtered states for state of one and two

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryMaterial.docx