Research Journal of Biological Science
MEDWELL PUBLICATIONS SINCE 2000

Using Time Series for Predicting the Occurrence of Hepatitis in Arar

[1, 2]Romany Mansour, [3]Khalid Buragga and [1]El-Saied Abdelrahim
[1]*Department of Computer Science, Faculty of Science, Northern Borders University, Arar, Saudi Arabia*
[2]*Department of Computer Science, Faculty of Science, N.V. Assiut University, Asyut, Egypt*
[3]*College of Computing and I.T., Northern Border University, Rafha, Saudi Arabia*

Key words: Epidemics, hepatitis, infectious disease control, preventable diseases, statistics

Corresponding Author:
Romany Mansour
Department of Computer Science, Faculty of Science, Northern Borders University, Arar, Saudi Arabia

Page No.: 60-63
Volume: 15, Issue 3, 2020
ISSN: 1815-8846
Research Journal of Biological Sciences

Abstract: Hepatitis is inflammation and one of the biggest public health problems and is one of the most important reasons that lead to the deaths in the world. In spite of efforts to treat the disease but there are many different cases of hepatitis type B and C. In the present study, we analyse data on the cases of hepatitis in Saudi Arabia, especially the Northern areas “Arar” and neighbouring cities and use this data to compare the prediction of disease models.

INTRODUCTION

Viral hepatitis, the liver inflammation and damage caused by a viral infection is a clear public health problem and is recognised as an important cause of morbidity and mortality worldwide[1]. Viral hepatitis is most commonly caused by one of five unrelated hepatotropic viruses (hepatitis A-C) which are the leading cause of hepatitis[1]. Clinical presentation can be either acute, as a result of a recent infection, or chronic[1].

The bulk of viral hepatitis cases are caused by hepatitis A-C and each virus has a distinct epidemiology and mode of transmission. Hepatitis A which causes in the region of 1.4 million infections per year, is transmitted via contaminated food and water as well as via. personal contact and 11-22% of cases require hospitalisation[2-4]. There is a strong socioeconomic component to infection, as access to clean drinking water is a key factor in prevention[2-4]. There is also an increased risk of developing hepatitis A associated symptoms such as liver failure with increasing age and this virus is a particular challenge in North Africa and the Middle East[2].

Hepatitis B, meanwhile, is spread sexually, although, perinatal/postnatal mother-to-child transmission is also an important route of contraction. In its early stages, it is highly contagious and it is thought that 350 million people are infected worldwide, leading to 1-2 million deaths per year[5, 6]. Although an effective vaccine has been available for >30 years, infections remain high, in part as it is only routinely given in the developed world[7].

Hepatitis C infection induces liver inflammation and fibrosis which ultimately lead to cirrhosis, however, infection with the virus can also cause hepatocellular carcinomas[8, 9]. There is currently no effective vaccine against hepatitis C which has been estimated to have a global infection prevalence of 2-3%[10, 11], although there is a large amount of regional variability with higher prevalence in Africa and the Middle East. Infection rates in countries such as Egypt, Cameroon, Saudi Arabia, Iraq and Syria can be as high as 15%[12-14].

In Saudi Arabia, hepatitis is an important health concern and although, cases of hepatitis A have dropped in recent years, those of hepatitis B and C remain high[15, 16]. In order to better understand the epidemiology of infection, disease surveillance data has been used previously to attempt to predict and prevent the spread of viral hepatitis. Efforts focussed on improvements in vaccination programmes, health advise and experimental therapies[17-20]. The types of longitudinal timescale data gathered from surveillance programmes are also amenable to mathematical modelling and time-series analysis and have been used to dissect the infection dynamics of other diseases. in order to better understand their spread and to predict future occurrences[21-24].

Modelling attempts to use infection timescale data to predict future disease trends. This may enable more effective provision and allocation of resources to improve public health. Such models, for example, might be able to anticipate an imminent epidemic or identify a specific region in which environmental factors are exacerbating infection. Although, complex modelling can be performed, e.g., artificial neural networks[18], simple modelling techniques can also be used to effectively determine data trends by extracting underlying patterns contained within the data series from the randomness. These 'decomposition methods' include, for example, the moving average method, exponential smoothing and linear regression. In each case, variations in the data such as seasonal changes and the long-term direction are used to forecast a likely future forecast that the data will take[25].

In the present study, we analyse data on incidence of hepatitis in the Kingdom of Saudi Arabia, particularly the Northern regions of Arar, Tarif, Rafha and Qurayyat and use this to develop a model with which future disease occurrence can be predicted. Two methods, exponential smoothing and the moving average method are discussed.

MATERIALS AND METHODS

Epidemiological data: Data were taken from hepatitis cases diagnosed between January 2005 and June 2012, in the Northern regions of the Kingdom of Saudi Arabia. Data were collated quarterly and patients classified by types.

Data analysis: Two techniques, the 'moving average' and 'exponential smoothing' methods were used to model the disease incidence data. For the averaging method, a periodicity of 3 (months) was used to calculate the rolling average. The seasonal component in the data was first removed and linear regression subsequently used to determine the overall trend in the data. Both the seasonal and trend components were then recombined in order to forecast the data into the future. For the exponential smoothing method, data were modelled using the formula:

$$F_{(t+1)} = \alpha A_1 + (1-\alpha) F_t$$

Where:

α = The smoothing factor

A and F = The values for actual and forecast cases, respectively

All modelling was performed using the Microsoft Excel Analysis Toolpak and margins of error quoted calculated from the descriptive statistics function.

RESULTS AND DISCUSSION

Data collected between January 2005 and June 2012 on the number of hepatitis cases in the Northern region of Saudi shown in Table 1. Data were stratified for kind A, B and C types in order to better understand disease dynamics. Case numbers were analysed along types A-C (Fig. 1). Although, there was a slight bias in cases towards B patients (~50%) compared to other types (~41 and ~8.5%). When comparing the age of patients presenting with hepatitis, there was a strong and significant bias towards the older age groups. There were no cases reported in anybody under the age of 30, suggesting that either these people were not susceptible to the virus or more likely, that the disease may remain

Table 1: Hepatitis case data from Northern regions used in this study

		Types		
Quarter	Cases	A	B	C
Jan.-Mar., 2005	50	36	14	0
Apr.-Jun., 2005	31	27	1	3
Jul.-Sep., 2005	72	41	21	10
Oct.-Dec., 2005	30	22	3	5
Jan.-Mar., 2006	30	23	4	3
Apr.-Jun., 2006	32	25	6	1
Jul.-Sep., 2006	54	43	6	5
Oct.-Dec., 2006	27	23	2	2
Jan.-Mar., 2007	25	24	1	0
Apr.-Jun., 2007	19	16	2	1
Jul.-Sep., 2007	16	6	9	1
Oct.-Dec., 2007	9	2	6	1
Jan.-Mar., 2008	108	41	65	2
Apr.-Jun., 2008	46	10	32	4
Jul.-Sep., 2008	23	8	15	0
Oct.-Dec., 2008	70	7	56	7
Jan.-Mar., 2009	26	1	22	3
Apr.-Jun., 2009	34	2	28	4
Jul.-Sep., 2009	19	5	12	2
Oct.-Dec., 2009	15	1	10	4
Jan.-Mar., 2010	19	5	12	2
Apr.-Jun., 2010	18	1	14	3
Jul.-Sep., 2010	12	1	11	0
Oct.-Dec., 2010	22	3	16	3
Jan.-Mar., 2011	21	0	17	4
Apr.-Jun., 2011	11	1	9	1
Jul.-Sep., 2011	9	0	9	0
Oct.-Dec., 2011	14	0	12	2
Jan.-Mar., 2012	26	1	22	3
Apr.-Jun., 2012	34	3	28	3
Total	922	378	465	79
Men.		0.41	0.5043	0.0857

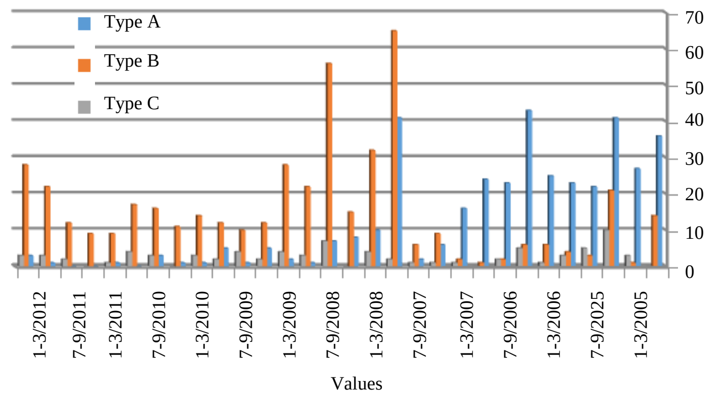

Fig. 1: Bar charts illustrating the breakdown of hepatitis cases in the Northern regions of Saudi Arabia by (A) type A, B and C

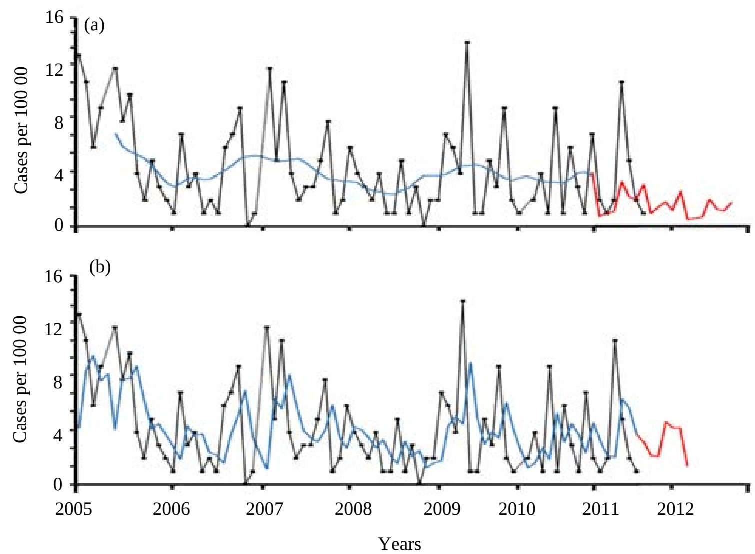

Fig. 2(a, b): Graphs illustrating the time course of hepatitis cases in this study (black), (a) Cases modelled using a moving average method (blue), also used to forecast future cases (red) and (b) Cases modelled using an exponential smoothing method (blue) and when used to forecast future cases (red)

undetected for many years and only becomes evident once significant damage has been done. This bias towards older patients is in line what has been seen in previous studies in the region[2, 26, 27].

Case data were plotted against time, in order to assess the dynamics of infection (Fig. 2). The data shows that month-by quarterly infections were quite erratic with many spikes in case numbers which make it difficult to model effectively (Fig. 2, black line). Both a running average method (Fig. 2a) and an exponential smoothing method (Fig. 2b) were used to try and forecast future disease incidence from the existing data. Using the rolling average method, both a seasonal variation and a general trend were identified from the data. The slight overall downward trend seen suggests that cases of hepatitis are dropping in the region over time. There have been reports that hepatitis B cases have been dropping in Arab countries[16], although, other types remain high currently.

Having separated the seasonal and erratic components from the data, it was possible to model the incidence of cases mathematically (Fig. 2a, blue line). The curve produced with this method effectively modelled the data, although, the Mean Absolute Percentage Error (MAPE) was high, at 11.5% (with an, a smoothing value of 0.5). This model could be used to provide a forecast the direction the data will take in future (Fig. 2 red line).

The exponential smoothing method (Fig. 2b) produced a model curve that more closely fitted the erratic nature of the data and gave an MAPE of 8.3%-slightly lower than the averaging method. Other studies that have compared the two methods have shown similar data to this study with respect to comparative performance[25]. In both models, the future incidence of hepatitis forecast remained relatively low, compared to historical levels, suggesting that current disease management programmes are being effective.

CONCLUSION

In conclusion, both running average and exponential smoothing can be used to provide an adequate model for Saudi hepatitis data, without the requirement for complicated computer analyses.

ACKNOWLEDGMENTS

This research was supported by the research project 435-036-5 “Predicting the incidence of hepatitis in Aara” financed by Deanship of Scientific Research Northern Border University and grate thanks for Saudi Epidemiology Bulletin to availability the data.

REFERENCES

01. WHO., 2014. Heath topics: Hepatitis. World Health Organization, Geneva, Switzerland.
02. Melhem, N.M., R. Talhouk, H. Rachidi and S. Ramia, 2014. Hepatitis A virus in the Middle East and North Africa region: A new challenge. J. Viral Hepatitis, 21: 605-615.
03. Matheny, S.C. and J.E. Kingery, 2012. Hepatitis A. Am. Family Physician, 86: 1027-1034.
04. Brundage, S.C. and A.N. Fitzpatrick, 2006. Hepatitis A. Am. Family Physician, 73: 2162-2168.
05. Liaw, Y.F. and C.M. Chu, 2009. Hepatitis B virus infection. Lancet, 373: 582-592.
06. Lai, M. and Y.F. Liaw, 2010. Chronic hepatitis B: past, present and future. Clin. Liver Dis., 14: 531-546.
07. Bertoletti, A. and C. Ferrari, 2012. Innate and adaptive immune responses in chronic hepatitis B virus infections: Towards restoration of immune control of viral infection. Gut, 61: 1754-1764.
08. El-Serag, H.B., 2012. Epidemiology of viral hepatitis and hepatocellular carcinoma. Gastroenterology, 142: 1264-1273.
09. Bartosch, B., R. Thimme, H.E. Blum and F. Zoulim, 2009. Hepatitis C virus-induced hepatocarcinogenesis. J. Hepatol., 51: 810-820.
10. Lavanchy, D., 2009. The global burden of hepatitis C. Liver Int., 29: 74-81.
11. Lavanchy, D., 2011. Evolving epidemiology of hepatitis C virus. Clin. Microbiol. Infect., 17: 107-115.
12. Guerra, J., M. Garenne, M.K. Mohamed and A. Fontanet, 2012. HCV burden of infection in Egypt: Results from a nationwide survey. J. Viral Hepatitis, 19: 560-567.
13. Nerrienet, E., R. Pouillot, G. Lachenal, R. Njouom and J. Mfoupouendoun *et al*., 2005. Hepatitis C virus infection in cameroon: A cohort-effect. J. Med. Virol., 76: 208-214.
14. Shepard, C.W., L. Finelli and M.J. Alter, 2005. Global epidemiology of hepatitis C virus infection. Lancet Infect. Dis., 5: 558-567.
15. Aljarbou, A.N., 2013. The emergent concern of hepatitis B globally with special attention to Kingdom of Saudi Arabia. Int. J. Health Sci., 7: 333-340.
16. Gasim, G.I., 2013. Hepatitis B virus in the Arab world: Where do we stand?. Arab J. Gastroenterol., 14: 35-43.
17. Goldstein, S.T., F. Zhou, S.C. Hadler, B.P. Bell, E.E. Mast and H.S. Margolis, 2005. A mathematical model to estimate global hepatitis B disease burden and vaccination impact. Int. J. Epidemiol., 34: 1329-1339.
18. Guan, P., D.S. Huang and B.S. Zhou, 2004. Forecasting model for the incidence of hepatitis A based on artificial neural network. World J. Gastroenterol. WJG., 10: 3579-3582.
19. Zou, L., W. Zhang and S. Ruan, 2010. Modeling the transmission dynamics and control of hepatitis B virus in China. J. Theor. Biol., 262: 330-338.
20. Mihigo, R., D. Nshimirimana, A. Hall, M. Kew, S. Wiersma and C.J. Clements, 2013. Control of viral hepatitis infection in Africa: Are we dreaming?. Vaccine, 31: 341-346.
21. Nobre, F.F., A.B.S. Monteiro, P.R. Telles and G.D. Williamson, 2001. Dynamic linear model and SARIMA: A comparison of their forecasting performance in epidemiology. Statistics Med., 20: 3051-3069.
22. Jose, M.V. and R.F. Bishop, 2003. Scaling properties and symmetrical patterns in the epidemiology of rotavirus infection. Philos. Trans. R. Soc. London Ser. B. Biol. Sci., 358: 1625-1641.
23. Benschop, J., M.A. Stevenson, J. Dahl, R.S. Morris and N.P. French, 2008. Temporal and longitudinal analysis of Danish swine Salmonellosis control programme data: Implications for surveillance. Epidemiol. Infect., 136: 1511-1520.
24. Armitage, P., G. Berry and J.N.S. Matthews, 2008. Statistical Methods in Medical Research. 4th Edn., John Wiley & Sons, Hoboken, New Jersey, Pages: 817.
25. Zhang, X., T. Zhang, A.A. Young and X. Li, 2014. Applications and comparisons of four time series models in epidemiological surveillance data. PLoS One, Vol. 9, No. 2. 10.1371/journal.pone.0088075
26. Ansaldi, F., A. Orsi, L. Sticchi, B. Bruzzone and G. Icardi, 2014. Hepatitis C virus in the new Era: Perspectives in epidemiology, prevention, diagnostics and predictors of response to therapy. World J. Gastroenterol. WJG., 20: 9633-9652.
27. Abdo, A.A., F.M. Sanai and F.Z. Al-Faleh, 2012. Epidemiology of viral hepatitis in Saudi Arabia: Are we off the hook? Saudi J. Gastroenterol., 18: 349-357.