JRHS

Journal of Research in Health Sciences

journal homepage: www.umsha.ac.ir/jrhs

Original Article

Real Time Detection of a Measles Outbreak using the Exponentially Weighted Moving Average: Does it Work?

Manoochehr Karami (MSc)[a], Hamid Soori (PhD)[b*], Yadollah Mehrabi (PhD)[c], Ali Akbar Haghdoost (PhD)[d], Mohammad Mehdi Gouya (MD)[e]

[a] *Department of Epidemiology, Faculty of Public Health, Shahid Beheshti University of Medical Sciences, Tehran, Iran*

[b] *Safety Promotion and Injury Prevention Research Center, Faculty of Public Health, Shahid Beheshti University of Medical Sciences, Tehran, Iran*

[c] *Department of Biostatistics, Faculty of Public Health, Shahid Beheshti University of Medical Sciences, Tehran, Iran*

[d] *Research Centre of Modeling in Health, School of Health, Kerman University of Medical Sciences, Kerman, Iran*

[e] *Centre for Disease Control, Ministry of Health & Medical Education, Tehran, Iran*

ARTICLE INFORMATION

* Correspondence

Hamid Soori (PhD)

Tel: +98 21 22431993

Fax: +98 21 22432036

E-mail: hsoori@yahoo.com

ABSTRACT

Background: There are few published studies that use real data testing to examine the performance of outbreak detection methods. The aim of this study was to determine the performance of the Exponentially Weighted Moving Average (EWMA) in real time detection of a local outbreak in Mashhad City, eastern Iran.

Methods: The EWMA algorithms (both $EWMA_1$ with $\lambda=0.3$ and $EWMA_2$ with $\lambda=0.6$) were applied to daily counts of suspected cases of measles to detect real outbreak which has occurred in the city of Mashhad during 2010. The performances of the EWMA algorithms were evaluated using a real data testing approach and reported by correlation analysis.

Results: Mashhad outbreak was detected with a delay of about 2 to 7 days using EWMA algorithms as outbreak detection method. Moreover, the utility of $EWMA_2$ algorithm in real time detection of the outbreak was better than $EWMA_1$ algorithm.

Conclusion: Applying the EWMA algorithm as an outbreak detection method might not be useful in timely detection of the local outbreaks.

Introduction

Measles outbreaks continue to affect people in the world[1]. In recent years, cases of outbreaks and sporadic events have been reported in Iran[2]. In the Iranian national surveillance system for communicable diseases, measles is a notifiable disease that must be reported as a clinical case when health care providers suspect it in a patient. Once reported, the surveillance process starts to work[3]. Suspected cases of measles is defined as individuals who have clinical findings including fever, generalized maculopapular rash and either a cough, coryza, or conjunctivitis. Such cases are reported to upper levels of the surveillance system on a daily bases[3].

Timely response to emerging diseases and outbreaks are a major public health and health systems priority. Due to limitations of traditional surveillance systems, especially in case of considering only confirmed cases of diseases, syndromic surveillance system has been implemented[4]. The most important feature of such systems is near real time detection of outbreaks. Generally, outbreak detection methods are under the umbrella of temporal and spatial methods[5] as main tools for syndromic surveillance systems. One of the most recognized methods/algorithms used by syndromic surveillance systems to detect outbreaks or any change in the disease trend is the Exponentially Weighted Moving Average (EWMA)[5-6]. EWMA algorithm is a statistical process control chart that effectively detects small and persistent changes of the disease trend[7]. However, there is no single algorithm like EWMA with constant parameter that can cover a wide range of outbreaks under different circumstances and settings.

There are three different approaches, which might be used by syndromic surveillance systems to examine the performances of outbreak detection algorithms including real data testing, synthetic simulation and semi-synthetic simulation[4]. Evaluating the performance of outbreak detection methods, using real data testing provide the highest degree of validity[8].

Consider both the strengths of this approach and the necessity of evaluation outbreak detection methods at different situations, this study was undertaken to determine the performance of the EWMA algorithm in real time detection of a local outbreak in the city of Mashhad, Iran, using the real data testing evaluation approach.

Methods

Data source for applying outbreak detection methods

Daily counts of suspected cases of measles in Mashhad City, eastern Iran including 56 cases reported to upper levels of the national surveillance system from 29 June to 4 August 2010 was considered as a prediagnostic data source. For timely detection of the Mashhad outbreak, we applied the outbreak detection method on this data source. The corresponding time series is shown in Figure 1.
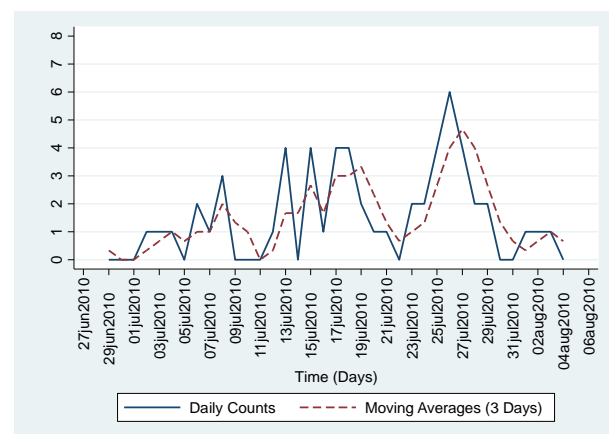
Figure1: Lines plot of daily counts of suspected cases of measles and corresponding moving averages (3 Days) in Mashhad from 29 June to 4 August 2010.

Data on the occurrence of a local outbreak in Mashhad including daily counts of confirmed cases of measles from 6 July to 28 July 2010 were obtained from provincial authorities of measles's surveillance system. During the outbreak days, 24 confirmed cases of measles were reported to the national surveillance system.

Outbreak detection method

The EWMA was used for detection of Mashhad outbreak. The EWMA monitors the trend of daily counts of suspected cases of measles. EWMA statistics is defined by the following recursive equation (1)[9]:

$$EWMA_t = \lambda Y_t + (1 - \lambda) EWMA_{t-1} \quad (1)$$

Where, Y_t equals number of suspected cases of measles in day t, λ is the weighting parameter that has been considered as 0.3 and 0.6 for $EWMA_1$ and $EWMA_2$, respectively. These weighting parameters have been estimated by authors using a fine- tuning approach and running a variety of the parameter, which ranges 0.1 to 0.9, and $EWMA_{t-1}$ is an estimated statistics at time t-1 i.e. one day before day t.

After calculating the $EWMA_t$ statistics, if its value is greater than the threshold level, the EWMA algorithm signals and warns the probable occurrence of an outbreak. Upper limit control or the threshold level of this statistics is calculated using the following equation (2):

Upper Control Limit=$EWMA_0 + k \times \sigma_{EWMA}$ (2)

Where, $EWMA_0$ is the mean of historical data at non outbreak times, the corresponding value according to non outbreak days is 1.5 cases and depicted as horizontal line in Figure 2 and 3, k is a constant parameter and considered as 2 in this work, and σ_{EWMA} is standard deviations of the estimated statistics of EWMA at times t to t_n.
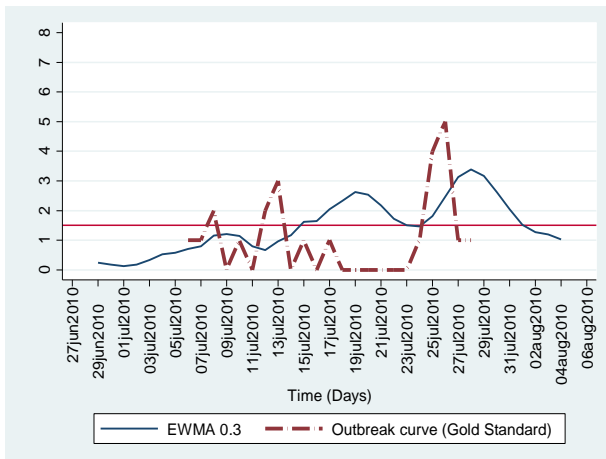
*The horizontal and red line shows the threshold level of EWMA algorithm which has been determined according to non outbreak days during 2010 in Mashhad.

Figure 2: Time series plot for estimated statistics of $EWMA_1$ versus the time series of the gold standard in Mashhad from 6 July to 28 July 2010.
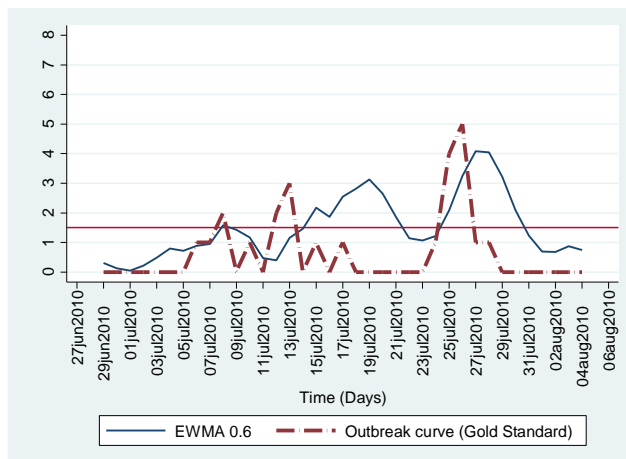
*The horizontal and red line shows the threshold level of EWMA algorithm which has been determined according to non outbreak days during 2010 in Mashhad.

Figure 3: Time series of estimated statistics of $EWMA_2$ and the time series of the gold standard in Mashhad from 6 July to 28 July 2010.

By considering the consequences of applying outbreak detection methods to the raw and syndromic data without removing the explainable patterns i.e. inability of the surveillance system's timely detection and response to the alarms[10], we preprocessed daily counts of suspected cases of measles using moving averages (3 days) and applied the EWMA algorithms on such smoothed data.

Evaluation approach for outbreak detection algorithm

Real data testing has been approached. We measured the performances of the EWMA algorithms for the timely detection of the Mashhad outbreak using correlation analysis approach[11]. Therefore, in this approach, we plotted the estimated time series statistics of suspected cases of measles ($EWMA_1$ and $EWMA_2$) alongside the time series of the gold standard, which had face validity and reflected actual outbreak activity. According to correlation analysis approach, if the time series for the estimated statistics of $EWMA_1$ exceeds the reference line or threshold level (Horizontal red line in related figures equals to 1.5 cases), an alarm is triggered. In other words, EWMA algorithm detects the interested outbreak once it exceeds the corresponding reference line. In addition, we used cross correlation function to find the time lag at which the correlation between the two time series (both $EWMA_1$ and $EWMA_2$ in relation to gold standard) is maximized as well as the strength of the correlation[12]. High correlation values between two the time series indicates good performance of the algorithm. We used Stata version 10 for data analysis.

Results

The line trend and corresponding smoothed graph of suspected cases of measles in Mashhad, reported to the national surveillance system between 29 June, and 4 August 2010, has been shown in Figure 1, using moving averages (3 Days). The mean and standard deviation of daily counts of suspected cases of measles during the mentioned period were 1.51±1.59. Figure 1 shows the increasing trend of suspected cases of measles at the beginning of the outbreak (6 July 2010). This trend declined once the outbreak ended on 28 July 2010 (Figure 1).

The lines trend regarding the estimated statistics of $EWMA_1$ algorithm after applying suspected cases of measles (weighing parameter=0.3) has been depicted in Figure 2. This

figure shows the time series of $EWMA_1$ versus the gold standard. This comparison exposes the weak performance of $EWMA_1$ concerning detecting the Mashhad outbreak in a timely manner. Our analysis reveals a seven-day lag in the prediction of outbreak using the $EWMA_1$ algorithm. In other words, there is about seven days lag between the start time of Mashhad outbreak (6 July 2010) and the time (13 July 2010) when the $EWMA_1$ algorithm exceeds the reference line.

We have found that applying $EWMA_2$ algorithm with a weighting parameter of 0.6 is superior in detecting the outbreak in a timely manner compared to the $EWMA_1$ output. The time series of $EWMA_2$ versus the gold standard have been shown in Figure 3 which indicates two days lag between the start time of Mashhad outbreak (6 July 2010) and the time (8 July 2010) when this algorithm exceeds the refer-

ence line. Consider to the start time of the interested outbreak i.e. 6 July 2010 and the 1st time while $EWMA_1$ and $EWMA_2$ statistics (which has been showed by dark blue color in Figure 2 and 3) exceeds the reference line (Red line in figure 2 and 3), we found different performance between $EWMA_1$ and $EWMA_2$ algorithms.

Figure 4 shows the cross correlation function of the time series of $EWMA_1$ and $EWMA_2$ in relation to the time series of the gold standard. The time series of $EWMA_1$ like $EWMA_2$ did not correlate well with the time series of the gold standard and precede Mashhad outbreak with delays. Peak correlation value for $EWMA_2$ was 0.60 at lag 2. The corresponding value for minimum correlation using cross correlation was 0.03 at lag 5.
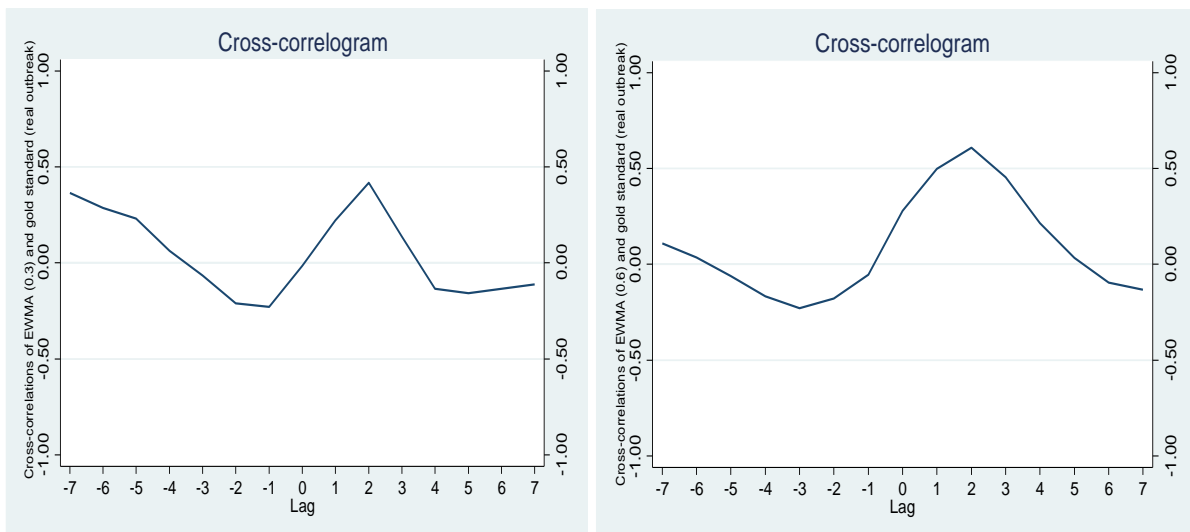
a.   Cross correlation function of the $EWMA_1$ algorithm

b. Cross correlation function of the $EWMA_2$ algorithm

Figure 4: Cross correlation functions of the time series of $EWMA_1$ (a) and $EWMA_2$ (b) in relation to the time series of the gold standard.

Discussion

Evaluating the performance of the outbreak detection methods using real data testing in this study assured its validity. This work indicated the weak performance of the EWMA algorithms in detecting a real outbreak. There are few published studies in literature, which evaluated the efficacy of such algorithms using real data testing approach[4, 6]. However, comparing the results of the present study is not confirmed by previously published studies because of

main differences including outbreak type, outbreak size and applied data source. For better clarification, we will discuss on the results of some previous studies before discussing the controversial issues surrounding these calculations.

Data on electrolyte sales, used by Hogan and colleague[13] for early detection of respiratory and diarrheal outbreaks during 1988 and 2001, found 90% correlation between electrolyte sales and hospital diagnoses. In addition, they

found that electrolyte sales detected outbreaks on average 2.4 weeks earlier than data on hospital diagnoses. As we have mentioned above, because of different characteristics of the data in this research and our work, the peak correlation values are not same or close. Results of a retrospective study, which aimed to predict respiratory and gastrointestinal outbreaks from chief complaints, indicated the efficacy of EWMA algorithm[14]. In another study[15], authors found good performance of the EWMA method in predicting the start and end of seasonal influenza in America.

Findings from similar studies[16-17], which implemented other evaluation approaches like synthetic and semi-synthetic simulation to determine the performance of EWMA algorithm in detecting simulated outbreaks, correspond with the above-mentioned studies. However, because of its different methodological approaches, it should not be compared with the results of the present study.

It seems that in addition to the role of characteristics of Mashhad outbreak and use of suspected case of measles as clinical data source in this study, there are two main reasons for convincing our findings. First, measles in Mashhad has had a low incidence rate. The second reason was originated from implementing elimination phase of measles disease in Iran and great interest of health system to detect any change in disease trend. For example, measles outbreaks have been defined as occurrence of four or more confirmed cases of measles[3]. Moreover, EWMA algorithm like other outbreak detection methods has its own strengths and limitations in different situations. According to the ideas of most researchers[8, 18], there is no single algorithm that can cover a wide range of outbreaks under different circumstances. Although, we preprocessed daily counts of suspected cases of measles using a moving average to considering the normality assumption, one of the reasons for poor performance of EWMA algorithms might be due to violation of the algorithm assumption.

The limitations of real data testing evaluation approach including the uncertainty about the size and start date of outbreaks and the inability to measure the performance of EWMA at

different circumstances are evident in this study.

Conclusion

We conclude that applying the EWMA algorithms, as an outbreak detection method at local levels is not suggested because of low incidence rate of measles and occurrence of small outbreaks. However, EWMA capabilities should be examined in detecting other outbreaks at different circumstances.

Acknowledgments

Conflict of interest statement

Authors have no conflicts of interest.

Funding

References

1. Monfort L, Munoz D, Trenchs V, Hernandez S, Garcia JJ, Aguilar AC, et al. Measles outbreak in Barcelona. Clinical and epidemiological characteristics. *Enferm Infecc Microbiol Clin*. 2010;28(2):82-86.

2. Esteghamati A, Gouya MM, Zahraei SM, Dadras MN, Rashidi A, Mahoney F. Progress in measles and rubella elimination in Iran. *Pediatr Infect Dis J*. 2007;26(12):1137-1141.

3. Zahraei SM, Dadras MN, Saborio A. *National guideline for measles surveillance (Elimination Phase)*. 1st ed. Tehran: Arvij; 2009. [Persian]

4. Chen H, Zeng D, Yan P. *Infectious Disease Informatics Syndromic Surveillance for Public Health and Bio Defense*. 1st ed. New York: Springer Science and Business Media; 2010.

5. Mandl KD, Overhage JM, Wagner MM, Lober WB, Sebastiani P, Mostashari F, et al. Implementing syndromic surveillance: a practical guide informed by the early experience. *J Am Med Inform Assoc*. 2004;11(2):141-150.

6. Buckeridge DL, Burkom H, Campbell M, Hogan WR, Moore AW. Algorithms for rapid outbreak

detection: a research synthesis. *J Biomed Inform*. 2005;38(2):99-113.

7. Zeng D, Chen H, Chavez C, lober WB, Thurmond MC. *Infectious Disease Informatics and Biosurveillance*. 1st ed. New York: Springer Science and Business Media; 2011.

8. Siegrist D, Pavlin J. Bio-ALIRT biosurveillance detection algorithm evaluation. *MMWR Morb Mortal Wkly Rep*. 2004;53 Suppl:152-158.

9. Lucas J, Saccucci M. Exponentially Weighted Moving Average Control Schemes: Properties and Enhancements. *Technometrics*. 1990;32:1-12.

10. Lotze TH, Murphy S, Shmueli G. Implementation and Comparison of Preprocessing Methods for Biosurveillance Data. *Advances in Disease Surveillance*. 2008;6:1-20.

11. Wagner MM. *Methods for evaluating surveillance data. In: Wagner MM, Moore AW, Aryel RM, eds. Handbook of Biosurveillance*. 1st ed. Oxford: Elsevier Academic Press; 2006. pp. 315-316.

12. Magruder SF. Evaluation of Over-the-Counter Pharmaceutical Sales as a Possible Early Warning Indicator of Human Disease. *J Hopkins Apl Tech D*. 2003;24(4):349-353.

13. Hogan WR, Tsui FC, Ivanov O, Gesteland PH, Grannis S, Overhage JM, et al. Detection of pediatric respiratory and diarrheal outbreaks from sales of over-the-counter electrolyte products. *J Am Med Inform Assoc*. 2003;10(6):555-562.

14. Ivanov O, Gesteland PH, Hogan W, Mundorff MB, Wagner MM. Detection of pediatric respiratory and gastrointestinal outbreaks from free-text chief complaints. *AMIA Annu Symp Proc*. 2003;2003:318-322.

15. Steiner S, Grant K, Coory M, Kelly H. Detecting the start of an influenza outbreak using exponentially weighted moving average charts. *BMC Med Inform Decis Mak*. 2010;10(1):37.

16. Wang XL, Wang QY, Liu DL, Zeng DJ, Cheng H, Li S, et al. Comparison between early outbreak detection models and simulated outbreaks of measles in Beijing. *Zhonghua Liu Xing Bing Xue Za Zhi*. 2009;30(2):159-162.

17. Jackson ML, Baer A, Painter I, Duchin J. A simulation study comparing aberration detection algorithms for syndromic surveillance. *BMC Med Inform Decis Mak*. 2007;7:6.

18. Aamodt G, Samuelsen SO, Skrondal A. A simulation study of three methods for detecting disease clusters. *Int J Health Geogr*. 2006;5:15.