http://www.jhpr.ir
Hosp Pract Res. 2023 March;8(1):212-216

doi 10.30491/hpr.2023.182070

HPR Hospital Practices and Research

Original Article

The Performance of Some Outbreak Detection Algorithms: Using the Reported COVID-19 cases in Iran

Mojtaba Sepandi [1], **Yousef Alimohamadi** [1*], **Mousa Imani** [1]

[1] Health Research Center, Life Style Institute, Baqiyatallah University of Medical Sciences, Tehran, Iran

***Corresponding Author:** Yousef Alimohamadi, Ph.D., Health Research Center, Life Style Institute, Baqiyatallah University of Medical Sciences, Tehran, Iran. Tel: + 98-9126386104, Email: y.alimohamadi67@gmail.com

Received January 8, 2023; **Accepted** February 7, 2023; **Online Published** March 15, 2023

Abstract

Background: Outbreak detection algorithms could play a key role in public health surveillance.

Objectives: This study aimed to compare the performance of three algorithms (EWMA, Cumulative Sum (CUSUM), and Poisson Regression) using the reported COVID-19 data for outbreak detection.

Methods: Three outbreak detection algorithms were applied to the data of COVID-19 daily new cases in Iran between 19/02/2020 and 20/06/2022, and 344 simulated outbreak days were injected into the data sequences. The Area Under the Receiver Operating Characteristics (ROC) Curve (AUC) and its 95% confidence intervals (95% CI) were also computed.

Results: EWMA9 had the lowest AUC (51%). Among the different algorithms, EWMA9 with $\lambda = 0.9$ and CUSUM 1 had the highest sensitivity with 100 and 87% (95% CI: 84%-91%), respectively.

Conclusion: According to the results, CUSUM, EWMA, and poison regression showed appropriate performance in detecting the COVID-19 outbreaks. These algorithms can be extremely helpful for health practitioners and policymakers in the detection of infectious disease outbreaks.

Keywords: Outbreak, Epidemic, Mathematics Concept, COVID-19, Iran

1. Background

Despite the development of preventive measures, COVID-19 remains a public health burden worldwide.[1] As of, September 26, 2022, 7,547,089 patients with COVID-19 have been identified in Iran, of which 144,394 deaths have occurred by the virus. According to statistics, Iran ranks 12th in the total number of deaths due to COVID-19.[2] An outbreak is defined as more cases of a disease than expected in a specific location over a specific period.[3] Outbreak detection algorithms could play a main role in effective public health surveillance.[4] Some techniques have been proposed and applied in practice for outbreak detection based on surveillance system data.[5] The exponentially Weighted Moving Average (EWMA) and Poisson Regression method are among the most known aberration detection algorithms.[6] The mentioned methods are based on a statistical process control approach to detect abnormalities in time series data.[7,8]

2. Objectives

This study compared the performance of three outbreak detection methods (EWMA, CUSUM, and Poisson Regression) using the reported COVID-19 data in the Islamic Republic of Iran.

3. Methods

First, data were collected on COVID-19 daily new cases in Iran between 19/02/2020 and 20/06/2022 through the Worldometer website available at: https://www.worldometers.info/coronavirus/. All registered cases during the mentioned period were included. Due to the lack of a gold standard to assess algorithm performance, a total of 344 simulated outbreak days were injected into the data sequences. In the next step, three outbreak detection algorithms were applied to the data.

3.1. EWMA

EWMA algorithm is defined by the following equation[9]:

$$EWMAt = Yt + (1\text{-}\lambda)\ EWMA\ t\text{-}1. \qquad (1)$$

Where Yt equals the number of suspected cases in day t, λ is the weighting parameter that has been considered as 0.1 for EWMA1, 0.2 for EWMA2, and so on. The upper control limit for outbreak detection is as follows:

$$\text{Upper Control Limit} = EWMA_0 + k \times \sigma_{EWMA}$$

Where k is a constant parameter, σ EWMA and EWMA0 are the standard deviation (σ) and the mean (μ) of data in the absence of the outbreak. In the current study, the amount of k was determined to be 2(K = 2), and the $\mu+2\sigma$ was considered as an upper limit for outbreak detection.

3.2. CUSUM
The CUSUM algorithm was used based on the following formula[10]:

$$CUSUM\ t = MAX\,(0, CUSUMt\text{-}1 + Yt - \sigma/2).$$

Where, Yt is the number of cases on day t (t = 1, 2... n), CUSUMt-1 is the value of CUSUM on day t-1 and σ is the standard deviation of the observed data on the non-outbreak days.

The upper control limit or level of alarm threshold for the CUSUM algorithm was calculated using the following equation:

$$\text{Upper Control Limit} = UCL = \mu + h \times \sigma$$

Where μ is the mean of the observed data on the non-outbreak days and h is an appropriate value (fixed parameter) ranging from 1 to 3 here. Also, σ is considered to be the standard deviation.

3.3. Poisson Regression Method
To determine the upper control limit, the expected cases per day were estimated as follows[6]:

$$Yt = \alpha i + \beta X$$

Where Yt is the expected cases at time t, and β is the coefficient of X, and X is the effective factors on the expected cases.

3.4. Statistical Analysis
Finally, the sensitivity, specificity, false alarm rate, and false negative rate as well as the negative and positive likelihood ratio of the three outbreak detection algorithms were computed. The AUC and its 95% confidence intervals (95% CI) were also computed. Greater values of AUC indicate better performance of a specific algorithm in comparison to other algorithms. All analyses were performed by Stata version 15 and Excel 2010.

4. Results
The overall sensitivity and specificity of the EWMA for all occurred outbreaks were 59% (95% CI: 53.6%-64.2%) and 0.89 (95% CI: 86.0%-92.0%) respectively. Overall sensitivity and specificity of the CUSUM for all occurred outbreaks were 72% (95% CI: 67.0%-76.7%) and 41.8% (95% CI: 37.5%-46.2%) respectively. In addition, the overall Sensitivity and Specificity of the Poisson Regression for all occurred outbreaks were 69% (95% CI: 64%-74%) and 85% (95% CI: 82%-89%) respectively. Among the different algorithms, EWMA9 with λ = 0.9 and CUSUM 1 had the highest sensitivity with 100 and 87% (95% CI: 84%-91%), respectively. EWMA9 and CUSUM 5 had the lowest specificity: 2% (95% CI: 1%-3%) and 23% (95% CI: 20%-27%), respectively (Table 1).

Table1. The Performance of Different Used Algorithms in the Detection of Outbreaks

Algorithm	Sensitivity	Specificity	False Alarm Rate	False Negative	LR+	LR-
EWMA1	0.53 (0.47-0.58)	0.99 (0.98-1.00)	0.01 (0.00-0.02)	0.47 (0.42-0.53)	44.64	0.48
EWMA2	0.54 (0.49-0.59)	1.00 (0.99-1.00)	0.00 (0.00-0.01)	0.46 (0.41-0.51)	136.87	0.46
EWMA3	0.54 (0.49-0.59)	1.00 (0.99-1.00)	0.00 (0.00-0.01)	0.46 (0.41-0.51)	275.22	0.46
EWMA4	0.54 (0.49-0.60)	1.00 (1.00-1.00)	0.00	0.46 (0.40-0.51)	-	0.46
EWMA5	0.54 (0.49-0.60)	1.00 (1.00-1.00)	0.00	0.46 (0.40-0.51)	-	0.46
EWMA6	0.54 (0.49-0.60)	1.00 1.00-1.00)	0.00	0.46 (0.40-0.51)	-	0.46
EWMA7	0.54 (0.49-0.60)	1.00 (1.00-1.00)	0.00	0.46 (0.40-0.51)	-	0.46
EWMA8	0.55 (0.49-0.60)	1.00 (1.00-1.00)	0.00	0.45 (0.40-0.51)	-	0.45
EWMA9	1.00 (1.00-1.00)	0.02 (0.01-0.03)	0.98 (0.97-0.99)	0.00	1.02	0.00
Cusum1	0.87 (0.84-0.91)	0.53 (0.49-0.57)	0.46 (0.42-0.50)	0.13 (0.09-0.16)	1.90	0.24
Cusum2	0.81 (0.77-0.86)	0.48 (0.44-0.52)	0.51 (0.46-0.55)	0.19 (0.14-0.23)	1.61	0.39
Cusum3	0.71 (0.66-0.75)	0.45 (0.40-0.49	0.54 (0.50-0.59)	0.29 (0.25-0.34)	1.30	0.66
Cusum4	0.63 (0.58-0.68)	0.40 (0.36-0.44)	0.59 (0.54-0.63)	0.37 (0.32-0.42)	1.08	0.91
Cusum5	0.58 (0.52-0.63)	0.23 (0.20-0.27)	0.76 (0.72-0.79)	0.42 (0.37-0.48)	0.76	1.83
Poisson Regression	0.69 (0.64-0.74)	0.85 (0.82-0.89)	0.15 (0.11-0.18)	0.31 (0.26-0.36)	4.72	0.37

The minimum amount of false alarm rate was related to EWMA2 to EWMA8, equal to 0%, and the maximum amount of false alarm was related to EWMA9, equal to 98% (95% CI: 97%-99%). The lowest false negative rate related to EWMA9 was equal to 0%. The highest value of the positive likelihood ratio was related to EWMA3 and was equal to 275.22, and the minimum negative likelihood ratio related to EWMA9 was equal to 0. The positive and negative likelihood ratio values have been shown in Table 1. Figure 3 shows the calculated AUC for different algorithms. EWMA9 had the lowest AUC (51%).

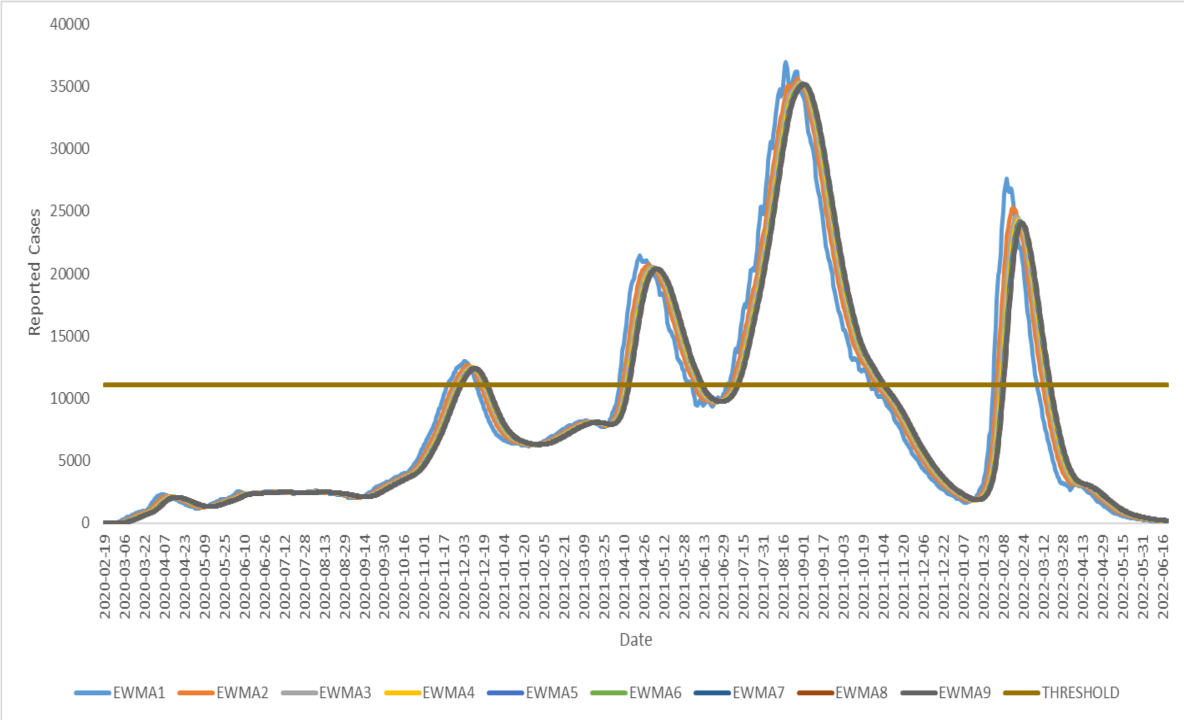

Figure 1. Line plot of reported COVID-19 cases from 19/02/2020 to 20/06/2022 and corresponding levels of overall alarm threshold according to EWMA.

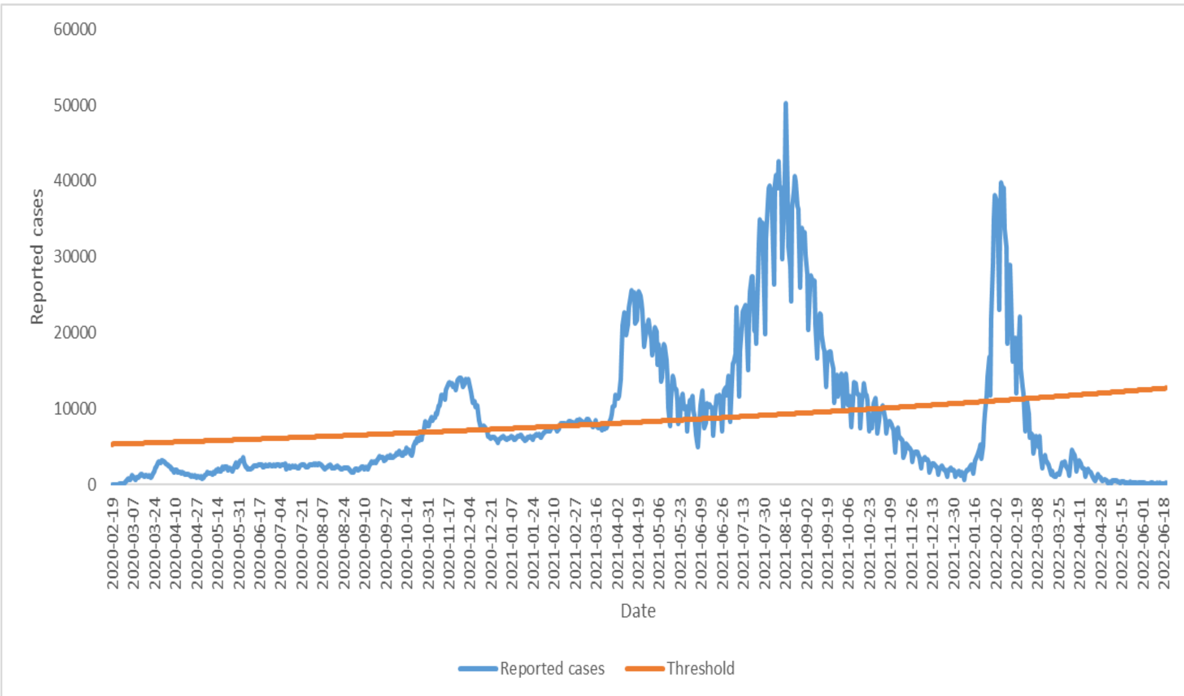

Figure 2. Line plot of reported COVID-19 cases from 19/02/2020 to 20/06/2022 and corresponding levels of overall alarm threshold according to Poisson regression algorithm.

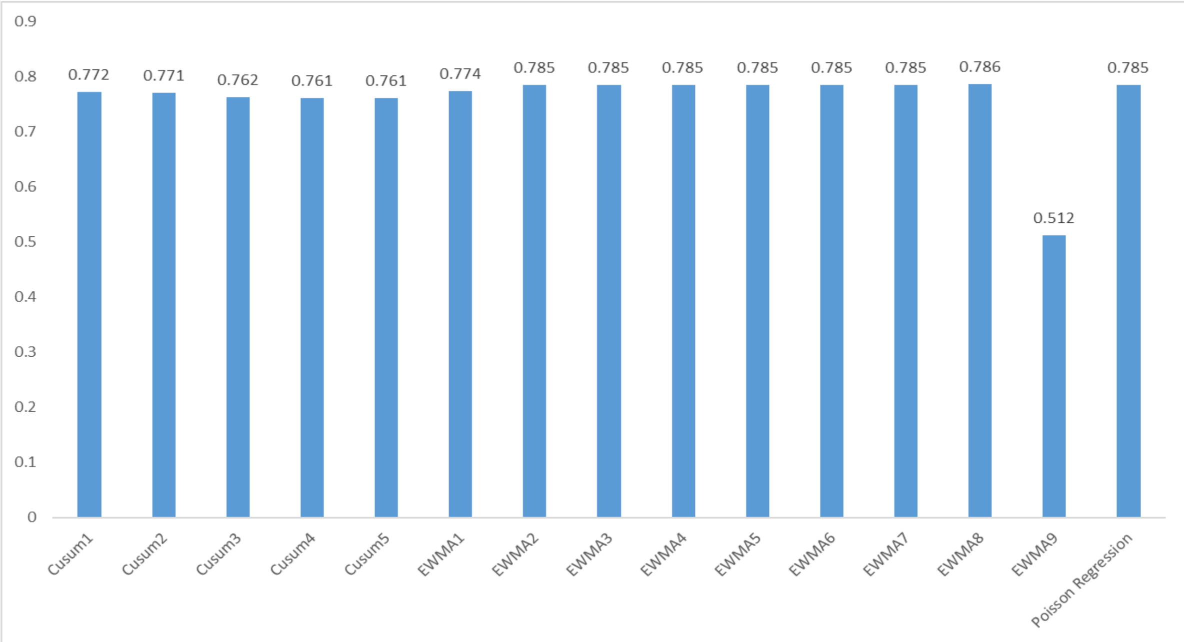

Figure 3. Area under the ROC Curve for Different Algorithms.

5. Discussion

The COVID-19 pandemic is a global public health challenge.[1] In Iran, cases of COVID-19 occur every day. Due to the nature of this infectious disease, controlling this epidemic is inevitable and necessary. Different methods have been used to determine the aberration from the normal trend of disease incidence in different studies.[3,11-13] This research analyzed the data of COVID-19 in Iran using CUSUM, EWMA, and Poisson Regression methods. Our results showed that among the different algorithms, EWMA9 with $\lambda = 0.9$ and CUSUM 1 respectively had the highest sensitivity, and EWMA9 and CUSUM 5 respectively had the lowest characteristics. Also, the lowest amount of false alarms was related to EWMA2 to EWMA8 and the highest amount of false alarms was related to EWMA9. In addition, the lowest false negative rate was related to EWMA9. The best positive likelihood ratio value was related to EWMA3. The lowest negative likelihood ratio was related to EWMA9 and this algorithm had the lowest AUC (51%). In the present study, the CUSUM algorithm was used to evaluate the spread of Covid-19 in Iran. This method has also been used in various studies in health data.[11-14] Our research focuses on the potential use of CUSUM to detect deviations in the trend of COVID-19. Due to its understanding and simplicity, the CUSUM algorithm can be useful for the early detection of deviations in the trend of COVID-19.[11]

The results of the present study are in line with the results of other studies[15,16] indicating the poor performance of the EWMA algorithm in diagnosing the spread of COVID-19 compared to Poisson Regression and CUSUM. However, some previously published studies report good performance for EWMA.[17,18] This discrepancy may be due to factors such as the type and size of the outbreak as well as the data sources. The performance of each model in outbreak detection depends on various factors such as the type of infectious diseases, the study location, the gold standard, and the accuracy of the records in the health care system.[19] Therefore, using this method alone is not recommended and is better as a combined method which can be used in the best way to detect the outbreak.[20-22] Furthermore, Poisson Regression was also used in this study to identify the prevalence of COVID-19. The performance of this method in detecting the spread of COVID-19 has also been investigated by previous studies.[11,23,24]

It is worth mentioning that just like any other study, the present research faced some limitations such as: the use of a simulated prevalence, which could differ from the actual prevalence. In addition, some cases of COVID-19 may not be accurately recorded in care systems.

Research Highlights

What Is Already Known?

Outbreak detection algorithms could play a key role in public health surveillance.

What Does This Study Add?

- According to the results, CUSUM, EWMA, and poison regression showed appropriate performance in detecting the COVID-19 outbreaks.
- These algorithms can be extremely helpful for health practitioners and policymakers in the detection of infectious disease outbreaks.
- It is not recommended to use a single method to detect outbreaks and it is better to use several methods together

6. Conclusion

According to the results, CUSUM, EWMA, and poison regression showed appropriate performance in detecting the COVID-19 outbreaks. So these algorithms can be extremely helpful for health practitioners and policymakers in the detection of infectious disease outbreaks. In general, it is not recommended to use a single method to detect outbreaks and it is better to use several methods together.

Acknowledgments

The authors are grateful for the cooperation of the management team of the hospitals in the data gathering process.

Author Contributions

MS and YA: Concept and design of the study, data collection, and manuscript drafting; MI: critical revision of the work.

Conflict of Interest Disclosures

All authors declared that they have no conflict of interest.

Ethical Approval

This study was approved by the Ethics Committee of Research at Baqiyatallah University of Medical Sciences, Iran (IR.BMSU.REC.1400.052).

References

1. Paakkari L, Okan O. COVID-19: health literacy is an underestimated problem. Lancet Public Health. 2020;5(5):e249-50. doi:10.1016/S2468-2667(20)30086-4
2. worldometers. [08 April 2020]; Available from: https://www.worldometers.info/coronavirus/#countries
3. Houlihan CF, Whitworth JA. Outbreak science: recent progress in the detection and response to outbreaks of infectious diseases. Clin Med. 2019;19(2):140-4. doi:10.7861/clinmedicine.19-2-140
4. Karami M. Validity of evaluation approaches for outbreak detection methods in syndromic surveillance systems. Iranian J Public Health. 2012;41(11):102-3.
5. Watkins RE, Eagleson S, Hall RG, Dailey L, Plant AJ. Approaches to the evaluation of outbreak detection methods. BMC Public Health. 2006;6(1):263. doi:10.1186/1471-2458-6-263
6. Alimohamadi Y, Zahraei SM, Karami M, Yaseri M, Lotfizad M, Holakouie-Naieni K. Alarm thresholds for pertussis outbreaks in Iran: National data analysis. Osong Public Health Res Perspect. 2020;11(5):309-18. doi:10.24171/j.phrp.2020.11.5.07
7. Shu L, Su Y, Jiang W, Tsui KL. A comparison of exponentially weighted moving average-based methods for monitoring increases in incidence rate with varying population size. IIE Trans. 2014;46(8):798-812. doi:10.1080/0740817X.2014.894805
8. Solgi M, Karami M, Poorolajal J. Timely detection of influenza outbreaks in Iran: Evaluating the performance of the exponentially weighted moving average. J Infect Public Health. 2018;11(3):389-92. doi:10.1016/j.jiph.2017.09.011
9. Lucas JM, Saccucci MS. Exponentially weighted moving average control schemes: properties and enhancements. Technometrics. 1990;32(1):1-12. doi:10.1080/00401706.1990.10484583
10. Karami M, Ghalandari M, Poorolajal J, Faradmal J. Early detection of meningitis outbreaks: Application of limited-baseline data. Iranian J Public Health. 2017;46(10):1366-73.
11. Alimohamadi Y, Mehri A, Janani M, Sepandi M. Aberration detection in influenza trends in Iran by using cumulative sum chart and period regression. J Taibah Univ Medical Sci. 2020;15(6):529-35. doi:10.1016/j.jtumed.2020.09.002
12. Bйdubourg G, Le Strat Y. Evaluation and comparison of statistical methods for early temporal detection of outbreaks: A simulation-based study. PloS One. 2017;12(7):e0181227. doi:10.1371/journal.pone.0181227
13. Dassanayake SK, French J. An Algorithm for Early Outbreak Detection in Multiple Data Streams. Online J Public Health Inform. 2019;11(1):e226. doi:10.5210/ojphi.v11i1.9942
14. Karami M, Soori H, Mehrabi Y, Haghdoost AA, Gouya MM, Esmailnasab N. Evaluating the performance of an outbreak-detection algorithms using semi-synthetic approach: cumulative sum algorithm. Iran J Epidemiol. 2013;9(2):29-38.
15. Karami M, Soori H, Mehrabi Y, Haghdoost AA, Gouya MM. Real time detection of a measles outbreak using the exponentially weighted moving average: does it work?. J Res Health Sci. 2012;12(1):25-30.
16. Bowen JR, Callander I, Richards R, Lindrea KB. Decreasing infection in neonatal intensive care units through quality improvement. Arch Dis Child Fetal Neonatal Ed. 2017;102 (1):F51-7. doi:10.1136/archdischild-2015-310165
17. 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;2003:318-22.
18. Steiner SH, Grant K, Coory M, Kelly HA. Detecting the start of an influenza outbreak using exponentially weighted moving average charts. BMC Med Inform Decis Mak. 2010;10(1):37. doi:10.1186/1472-6947-10-37
19. Spreco A. Epidemiological and statistical basis for detection and prediction of influenza epidemics. Doctoral dissertation, Linköping University Electronic Press; 2017.
20. Texier G, Allodji RS, Diop L, Meynard JB, Pellegrin L, Chaudet H. Using decision fusion methods to improve outbreak detection in disease surveillance. BMC Med Inform Decis Mak. 2019;19(1):38. doi:10.1186/s12911-019-0774-3
21. Dawson P, Gailis R, Meehan A. Detecting disease outbreaks using a combined Bayesian network and particle filter approach. J Theor Biol. 2015;370:171-83. doi:10.1016/j.jtbi.2015.01.023
22. Faryadres M, Karami M, MOGHIM BA, Esmailnasab N, Pazhouhi K. Levels of alarm thresholds of meningitis outbreaks in Hamadan Province, west of Iran. J Res Health Sci. 2015;15(1):62-5.
23. Souty C, Jreich R, Le Strat Y, Pelat C, Boëlle PY, Guerrisi C, et al. Performances of statistical methods for the detection of seasonal influenza epidemics using a consensus-based gold standard. Epidemiol Infect. 2018;146(2):168-76. doi:10.1017/S095026881700276X
24. Toubiana L, Flahault A. A space-time criterion for early detection of epidemics of influenza-like-illness. Eur J Epidemiol. 1998;14:465-70. doi:10.1023/A:1007481929237